

Complete GFW Rulebook for Wikipedia

Xia Chu*

Version 2.0, October 18, 2013[†]

1 Abstract

In this report, we detail the *complete* and *exact* rulebook that the Great Firewall of China (GFW) exerts on Wikipedia. We call it “rulebook” (instead of the common term “blacklist”) because we not only identify the blacklisted terms, but also the exact string matching rules deployed by GFW. An efficient probing methodology makes this possible.

GFW blocked Wikipedia outright in the early years but gradually loosened the blockage, first by unblocking all non-Chinese versions, then by unblocking the Chinese version, except for certain entries deemed harmful by the Chinese authority.

There have been some efforts in understanding the Wikipedia blacklist, for example, at time of writing, the site Greatfire.org [1] tracks ~ 700 Wikipedia pages and ~ 400 are claimed to be blocked or partially blocked in China.

Wikipedia contains millions of pages, e.g. more than 700,000 articles for the Chinese version, and more than 4,240,000 articles for the English version. It seems a daunting and unfeasible task to test these pages exhaustively, hence there has been no well known attempt to gather the complete blacklist.

While a small sample of the blacklist is useful, the complete picture can be much more powerful in revealing the underlying works of GFW and its operators. In this study, we devised a methodology which efficiently examines the entire Wikipedia corpus, hence exposing to the world the complete GFW rulebook for Wikipedia the first time. In total, there are 919 rules (excluding URL terms) which are applicable to Wikipedia, affecting 5336 pages in Chinese Wikipedia and 67 English Wikipedia pages.

The revealed rulebook also demonstrates that the GFW operation is haphazard and ill-maintained. At the same time, Chinese censorship bureaucracy *intends* to be thorough and extensive.

To be precise, the findings in this report are on two Wikipedia snapshots: 2013-09-08 for the Chinese version and 2013-09-04 for the English version.

In Version 2.0, we studied GFW’s filtering rules for HTTP responses extensively and identified a comprehensive list (including those affecting Wikipedia and beyond). This list is small (19 items) but they affect many more pages on Wikipedia and other websites.

1.1 Structure of the Report

- Section 1: The abstract.
- Section 2: The background on GFW and keyword-based filtering.
- Section 3: The methodology (for HTTP requests).
- Section 4: Four types of GFW rules for Wikipedia.
- Section 5: GFW filtering rules for HTTP responses and the methodology (**new in Version 2.0**).
- Section 6: Wikipedia specifics, e.g. its content structure and certain features that are relevant to blockage.
- Section 7: Caveats and cautions when interpreting the list.
- Section 8: The complete GFW rulebook for Wikipedia.
- Section 9: Conclusion remarks.
- Appendix A: Diagnosis of Greatfire.org’s Wikipedia list.
- Appendix B: Observations of self censorship attempts on Chinese Wikipedia.

*Author contact: Email: summer.agony@gmail.com; Twitter: @SummerAgony.

[†]Version 1.0 was released on October 1, 2013.

2 Background

2.1 Related Work

There are several attempts in understanding various blacklists employed by China's censors.

- Greatfire.org [1] is a website dedicated to track China censorship, including websites (URLs and IPs), Weibo search blacklist, blocked search terms on search engines and Wikipedia pages. Their Wikipedia list is very helpful to our research, and we now have a full diagnosis of that list (see Appendix A).
- Citizen Lab operates the site china-chats.net [2], based on the paper by Crandall et al. [3], which reverse-engineered the blacklists employed by two popular IM clients in China.
- China Digital Times [4] operates a crowd-sourcing project to track the search blacklist applied on Weibo.com.
- ConceptDoppler, a 2007 paper by Crandall, et al [5], covers GFW and its keyword-based filtering in great details. Furthermore, it developed a smart methodology based on LSA (latent semantic analysis) to pick potentially sensitive terms for probing and identified 122 blacklisted keywords. That list is apparently outdated, but their probing methodology remains valid, and is the basis of the methodology we use in this report.

The author hopes that the information revealed in this report significantly increase the “transparency” of GFW.

2.2 GFW and Keyword-based Filtering

The technical details of GFW have been studied fairly thoroughly since more than 10 years ago [6] [7]. In summary, GFW uses a combination of IP blocking, port blocking, DNS poisoning, and DPI-based IDS (intrusion detection system), etc, to disrupt normal internet traffic.

Our focus here is the keyword-based filtering, which is been used more and more widely by Chinese censors due to its precision, flexibility and scalability. In a nutshell, GFW devices apply pattern matching on internet traffic, when it finds a match, it sends out forged TCP RESET packets to the server and the client to interrupt the TCP connection. Worse off, the reset stays effective for 90 seconds, leaving the user unable to connect to the server in this duration. Note that not many users know about the 90 seconds interval, and they usually give up and “learn” that the website is “unstable”. Certain people believe that the site would remain blocked “from several minutes to up to an hour”, but I have not seen documented evidences for these claims.

Actually when GFW blocks a website, keyword-based filtering is almost always used, either on its own or in combination with DNS poisoning and/or IP blocking. This is because DNS poisoning can be circumvented by using alternative DNS servers or by editing the *host* file (the actual filename and path depends on operation systems), and IP blocking requires “maintenance” work to diligently track IP address changes of these sites. The keyword-based approach is a low-maintenance and high-efficiency one. From my experience, GFW uses keyword-based blocking for almost all blocked sites, DNS poisoning is used for a subset, and IP blocking is reserved only for those high-impact, big-name websites.

2.3 HTTP Request Scan and HTTP Response Scan

There are two types of keyword-based filtering. One type is applied to HTTP requests, the other type is applied to HTTP responses. It is very important to distinguish between the two, because they are two separated systems. The filtering rules are entirely different, as well as the user experience.

An HTTP request is sent by the user's browser to web server for content, it is well structured and much smaller than HTTP responses both in size and quantity. GFW's filtering rules for HTTP requests are about 1600 (excluding website URLs), and they contain all kinds of stuff, from very general terms to arcane strings. We identified a *complete* set of 919 rules (excluding website URLs) that affect Wikipedia. On the contrary, GFW's filtering rules for HTTP responses are much more specific. Our extensive study identified only 19 of them. Even though we can not claim this small list is *complete*, we have high confidence that we have not missed many. This drastic difference is because scanning HTTP responses is way more complicated and costly than scanning HTTP requests.

These two types of filtering render different user experience. If a user visits a page which offends GFW in its HTTP request scan, he will almost immediately get a “connection reset” error in the browser. If the page URL passes GFW's

HTTP request scan but the page content contains something offends GFW's HTTP response scan, he will often see a partially loaded page which just hangs there.

The blocking rates for these two types are very different as well. For HTTP request scan, in my hundreds of thousands of tests, GFW reset more than 99% of the offending HTTP requests. For HTTP response scan, the reset rates have wide variation among different tested IPs and rules, and they are usually in the range of 60% to 95%.

GFW's filtering for HTTP requests is much easier to study and indeed studied much more than that for HTTP responses. In particular, **Version 1.0** of this report is on GFW's filtering rules for HTTP requests, not HTTP responses. **Version 2.0** adds the learnings on GFW's HTTP response filtering.

2.4 HTTP Request Scan Details

We examined how exactly GFW scans HTTP request. For illustration, let's look at the following HTTP request to the page http://en.wikipedia.org/wiki/Xu_Zhiyong as an example:

```
GET /wiki/Xu_Zhiyong HTTP/1.1\r\n
Host: en.wikipedia.org\r\n
Connection: keep-alive\r\n
Referer: http://en.wikipedia.org/wiki/Main_Page\r\n\r\n
User-Agent: [omitted]\r\n
[other headers omitted]
```

From extensive study, we found that GFW pulls out the *Host* field (`en.wikipedia.org` in the above example) and the *GET* target (`/wiki/Xu_Zhiyong HTTP/1.1` in the above example) and concatenates the two, i.e. `en.wikipedia.org/wiki/Xu_Zhiyong HTTP/1.1`, then GFW applies its rulebook to this string and checks if it matches any. GFW does not seem to care about other HTTP request types like POST, HEAD etc, these other request types are not for retrieving page content.

A common (and quite natural) hypothesis is that GFW checks the Host field to see if it is Wikipedia and checks the GET target to see if it contains a sensitive term. We tested it extensively and found no indication of this, or any other way GFW handles the Host field and the GET target other than as described in the previous paragraph. This is understandable from an implementation efficiency perspective. My theory is that GFW has one component to do the *extraction* (i.e. identifying HTTP request, pulling out Host and GET fields and concatenating), and a separate component to do the *pattern matching*.

2.5 GFW and Wikipedia

For Wikipedia, GFW blocked the entire site on June 3, 2004. There were temporary lifts for several times after that. After a meeting between Wikipedia founder Jimmy Wales and China authority in September 2008, the site became mostly accessible. China authority changed to use a keyword-based approach to block access to certain individual pages, which is the subject of this study.

Wikipedia does offer an HTTPS version which bypasses GFW's keyword-based filtering, but the HTTPS version is usually blocked via a host-port block. The most recent block started on May 31, 2013 and remains on, or partially on.

3 Methodology (HTTP Request Scan)

This section is about the methodology to study GFW's filtering rules for HTTP requests. Section 5 covers the methodology to study GFW's HTTP response filtering and the learnings.

3.1 Wikipedia Dump

Wikipedia offers content dump, from which we can gather all entries in the current or previous snapshots. In this study, we downloaded the dump for Chinese and English on Sep 8, 2013, from <http://dumps.wikimedia.org/zhwiki/20130908/> and <http://dumps.wikimedia.org/enwiki/20130904/> respectively.

Field	Type	Null	Key	Default	Extra
page_id	int(10) unsigned	NO	PRI	NULL	auto_increment
page_namespace	int(11)	NO	MUL	NULL	
page_title	varbinary(255)	NO		NULL	
page_restrictions	tinyblob	NO		NULL	
page_counter	bigint(20) unsigned	NO		0	
page_is_redirect	tinyint(3) unsigned	NO	MUL	0	
page_is_new	tinyint(3) unsigned	NO		0	
page_random	double unsigned	NO	MUL	NULL	
page_touched	binary(14)	NO		&	
page_latest	int(10) unsigned	NO		NULL	
page_len	int(10) unsigned	NO	MUL	NULL	

Table 3.1: Database schema for the page table of Wikipedia

Information for the Wikipedia entries is offered as a MySQL database dump to be downloaded. We mainly work with the `page` table, whose schema is shown in Table 3.1.

Each record in the `page` table corresponds to one Wikipedia page, whose URL can be constructed from `page_namespace` and `page_title`. The following are two examples (more details in Table 6.1).

- Record [page_id=221324; page_namespace=0; page_title=“Princelings”] corresponds to the URL: <http://en.wikipedia.org/wiki/Princelings> (an article).
- Record [page_id=2791496; page_namespace=1; page_title=“Princelings”] corresponds to the URL: <http://en.wikipedia.org/wiki/Talk:Princelings> (the Talk page for that article).

We did not test all entries in this table, since some of these entries are insignificant or not applicable. To be precise, for the Chinese version, we tested all entries except for MediaWiki (namespace=8) and MediaWiki talk (namespace=9); for the English version, we tested all entries in Article (namespace=0), User (namespace=2), Project (namespace=4), File (namespace=6), Template (namespace=10), Help (namespace=12), and Category (namespace=14). In total, we examined 3,078,365 entries for the Chinese version and 20,631,910 entries for the English version. The actual number of examined Chinese Wikipedia pages is one order of magnitude bigger, because there are 9 variants of the Chinese Wikipedia (see Section 6.2). Section 6 contains more details about Wikipedia content.

3.2 Probing

Our probing is based on the probing methodology described in the ConceptDoppler paper [5]. Basically we first establish a valid TCP connection to a destination across GFW, then we send probing TCP packets to observe response. For more details, please read that well-written paper which is freely available online.

Our main improvements are in the following three areas:

- Firstly, in one probing request, we can put in multiple phrases for testing. As we have described in Section 2.2, to test if a Wikipedia URL is on GFW rulebook, we can put the entire URL (or multiple such URLs) in the GET target of the HTTP request for testing. This dramatically increases the probing efficiency, since majority of the Wikipedia pages do not trigger GFW reset.

There is a limit on the size of the probing request. When the request is too long, it may get truncated or the destination host may return various errors which may “pollute” the signal we are looking to capture. This size limit seem to vary among different destination hosts. So it is conservative to not test too many phrases in one run, and it is good practice to verify the size is appropriate for a destination host, e.g. examining false positive rate and false negative rate by sending known offending requests or known non-offending requests with varying size.

- Secondly, like we have described in Section 2.2, we can perform the test on any web server as long as it’s on the other side of GFW. It does not need to be a Wikipedia server. Furthermore, when GFW sees an offending packet, it only resets the connection between the client IP and the server IP and blocks it for 90 seconds. Connection

between the user IP and other hosts are not affected. So we can run many tests simultaneously. In fact, we ran dozens of probing threads concurrently in this study. This is the crucial reason why we can examine millions of Wikipedia URLs using just one machine in a short period.

- Lastly, we built a pipeline to streamline the entire probing process. For example, if one probing request containing 20 Wikipedia URLs triggers reset, the program would use an efficient search algorithm to find which URL(s) is the culprit. Furthermore, given one “offending” URL, we wrote a program to automatically examines what is the actual GFW string matching rule that it offends. Section 3.3 is devoted to this aspect.

With these improvements, our program can examine tens of millions of URLs in a reasonable time frame.

That said, there are many nitty-gritty issues with GFW and Wikipedia, thus significant babysitting and manual inspection is still required. But that is manageable and not more than a few days of work for an individual. More importantly, these manual inspection and investigation greatly improved the author’s understanding of GFW, and Wikipedia as well.

3.3 Identifying GFW String Matching Rules

It is well known that identifying regular expressions with no constraints is costly and difficult. In the GFW case, we make one assumption which makes it much easier. GFW need to be efficient, so it is reasonable to assume that it does not use unnecessarily complex regular expressions. Indeed, hundreds of cases I’ve examined show that GFW only uses simple plain string matching like the following:

- Single: $\text{\$target_string} \sim \text{\$term}$
- Double: $\text{\$target_string} \sim \text{\$term_1} \ \&\& \ \text{\$target_string} \sim \text{\$term_2}$
- Triple: $\text{\$target_string} \sim \text{\$term_1} \ \&\& \ \text{\$target_string} \sim \text{\$term_2} \ \&\& \ \text{\$target_string} \sim \text{\$term_3}$
- or more

In all the cases I’ve examined, GFW never uses regex features like metacharacter, character classes, or boolean “or” (boolean “or” can be viewed as equivalent to multiple rules). This is not unexpected, because GFW devices need to scan the huge internet traffic between China and the world in real time, for this purpose, plain string matching combined with boolean “and” is the most cost-effective approach.

However, GFW is case-insensitive for all cases tested.

Once we establish this assumption, we can very quickly identify the GFW rule given a string (Wikipedia page URL in this case) which triggers GFW reset. We will explain the general procedure below. To help with explanation, we use the following notations:

- S is a string.
- $L(S)$ is the string S with its leftmost character removed.
- $R(S)$ is the string S with its rightmost character removed.
- $\text{concat}(S1, S2)$ is the concatenation of two strings $S1$ and $S2$. Note that in order to avoid odd cases where the concatenation introduces new sensitive terms, we actually add two whitespaces between $S1$ and $S2$ here.
- $M(S)$ is short for $\text{concat}(L(S), R(S))$.

For example, the string “en.wikipedia.org/wiki/Princelings” (S) triggers GFW reset, we then check the strings “n.wikipedia.org/wiki/Princelings” ($L(S)$) and “en.wikipedia.org/wiki/Princeling” ($R(S)$) to see if they trigger reset. If neither triggers, we check $M(S)$, if that still does not trigger, then we are certain that the string S is on GFW rulebook. In this (quite special) example, both $L(S)$ and $R(S)$ trigger reset, so we need to strip more characters from the beginning and the end, and repeat the process. The actual GFW rule turns out to be “.wikipedia.org/wiki/Princeling” (T). For a string T on GFW rulebook that does not involve boolean “and”, the necessary and sufficient condition (under the aforementioned assumption) is that T triggers reset but $M(T)$ does not, for example, $M(T)$ is “wikipedia.org/wiki/Princeling .wikipedia.org/wiki/Princelin” here.

For other cases where S triggers reset, $R(S)$ and $L(S)$ do not but $M(S)$ does, this means GFW is using boolean “and” here, i.e. GFW’s rule here is “ $T1 \ \&\& \ T2$ ” where $T1$ is a substring of $R(S)$ and $T2$ is a substring of $L(S)$. In this case, we hold $R(S)$ unchanged while repeating the procedure on $L(S)$, until we find a substring of $L(S)$ (denoted as $T1$), where $\text{concat}(T1, R(S))$ triggers reset, but $\text{concat}(M(T1), R(S))$ does not. Then we hold $T1$ and repeat this procedure for $R(S)$, until we find a substring of $R(S)$ (denoted as $T2$), where $\text{concat}(T1, T2)$ triggers reset, but $\text{concat}(T1, M(T2))$ does not. Then we recheck $\text{concat}(M(T1), T2)$, if that does not trigger rest, we are certain that the GFW rule is “ $T1 \ \&\& \ T2$ ”.

This procedure also applies to “and” rules with more than two parts. For all Wikipedia-related cases, we have not observed any rules involving more than two parts. Also, for rules like “ $T1 \&\& T2$ ”, we examined whether GFW might be using regex rule like “ $T1.*T2$ ”, but that never happens.

This procedure may seem to require many steps, luckily for majority of cases, the GFW rule is standard, so it can be a lot more efficient. For a given string S that triggers reset, we first check $M(S)$. If it is fine, then we are done. If it is not, then we suspect it is two strings with boolean “and”. We hypothesize that $T1$ is the domain part (e.g. “zh.wikipedia.org”) and $T2$ is the page title, so we check $concat(M(T1), T2)$, $concat(T1, M(T2))$ and $concat(T1, T2)$, if the first two are fine and the last triggers reset, then our hypothesis holds and we are done. If this is not the case, then we need to do more testing. For the entire study, there are less than 20 non-standard cases (i.e. “corner cases” as in Table 4.2).

Note that these testings may give false positives or false negatives. So we need to do such testing multiple times, and for the corner cases, we conduct dozens of tests to verify it is indeed a nonstandard rule. Furthermore, after we identify a rule, we will also examine its traditional or simplified versions, even when those version do not have a Wikipedia page. With all these rigorous tests, the author claim that all rules reported in this document are accurate and precise with full confidence.

4 GFW Rules for Wikipedia

The keyword-based filtering part of GFW can be viewed as a collection of string matching rules. This collection is applied to all HTTP requests that pass through a GFW device. Understandably, some (or more precisely, many) of these are not targeting Wikipedia, which is the focus of this study. So we only cover those rules affecting Wikipedia.

We can group these into four types (Table 4.1).

Rule Type	Affects	Count	Example	String Matching Rule for the Example
Broad (targeting Wikipedia)	any Wikipedia page whose title contains the string	324	俄羅斯	俄羅斯 & zh.wikipedia.org
Prefix (targeting Wikipedia)	any Wikipedia page whose title starts with the string	577	中華民國總統 Charter_08	zh.wikipedia.org/wiki/中華民國總統 en.wikipedia.org/wiki/Charter_08
SELF (Non-URL)	all internet traffic, including Wikipedia	18	延安日记	延安日记
URL	all internet traffic, including Wikipedia	37	blogspot.com	blogspot.com

Table 4.1: Four types of GFW rules affecting Wikipedia

- $\$term \& zh.wikipedia.org$. In this case, if $\$term$ appears anywhere in the Wikipedia page’s title, the article is blocked (for that version). We’ll refer to this type as “*broad*” match. There are 324 rules of this type. Although most of such rules affect only one page, some of these affects many more, e.g. 俄羅斯¹ & zh.wikipedia.org affects 688 pages (284 articles and 404 other pages) and 1989年² & zh.wikipedia.org affects 232 pages (105 articles and 127 other pages).

A notable phenomenon is that there is currently not a single rule like $\$term \& en.wikipedia.org$, i.e., all broad rules target the Chinese Wikipedia. All but 8 of the broad rules are like $\$term \& zh.wikipedia.org$. Table 4.2 contains these corner cases where the Wikipedia part is not zh.wikipedia.org and the actual terms.

- zh.wikipedia.org/wiki/ $\$term$ or en.wikipedia.org/wiki/ $\$term$. In this case, if the article’s title starts with $\$term$, it is blocked. We’ll refer to this type as “*prefix*”. There are 577 rules of this type. For most of the time there is only one article that matches the rule when the article’s title equals $\$term$, but there

¹俄羅斯: Russia, in traditional Chinese.

²1989年: Year 1989. This is the year of the Tiananmen Protest, a.k.a, June 4th Movement.

are cases when one rule affects many many articles, for example, the rule `zh.wikipedia.org/wiki/中華民國總統`³ affects 18 articles.

Note that Chinese Wikipedia has 9 variants (see Section 6.2), so the `zh.wikipedia.org/wiki/` part could be in other form (e.g. `zh.wikipedia.org/zh-cn/`). There are 4 corner cases where the wikipedia part is not `zh.wikipedia.org/$variant/` or `en.wikipedia.org/wiki/`. See Table 4.2.

- `$term` where `$term` is not a URL. In this case, *any* HTTP GET request that matches `$term` will trigger reset, for example, when one searches for the term on a search engine across GFW, the connection will be blocked. In our case, if a Wikipedia page's title is `$term`, or contains `$term`, it will be blocked. For example, 延安日记⁴ by itself is a GFW rule, so the Wikipedia page `http://zh.wikipedia.org/wiki/延安日记` is blocked. This type is relatively rare, only 18 of them affect Wikipedia.
- `$term` where `$term` is a URL. These rules are intended for blocking these sites, and there are consequences for Wikipedia if Wikipedia has such an article. For example `blogspot.com` is on GFW rulebook, so the Wikipedia page `http://en.wikipedia.org/wiki/Blogspot.com` is blocked. GFW rulebook contains a large number of URLs, and many Wikipedia pages are affected. There are 7 terms that affect a Chinese Wikipedia page (all pages considered) and 30 terms that affect an English Wikipedia article (i.e. namespace=0). There are many more that affect English non-article pages, which are omitted from this report.

Rule Type	String Matching Rule	Count	Term
broad	zh.wikipedia.org & \$term	316	
<i>broad corner cases:</i>			
broad	wikipedia & \$term	2	零八宪章, 零八憲章
broad	wikipedia.org & \$term	3	万里, 陆肆事件, 陸肆事件
broad	zh.wikipedia & \$term	1	香港市民支援愛國民主運動聯合會
broad	zh.wikipedia.org/w & \$term	1	历史的伤口
broad	zh.wikipedia.org/zh-hk & \$term	1	5YWt5Zub5LqL5Lu2
prefix	zh.wikipedia.org/wiki/\$term	259	
prefix	zh.wikipedia.org/zh/\$term	35	
prefix	zh.wikipedia.org/zh-cn/\$term	85	
prefix	zh.wikipedia.org/zh-tw/\$term	51	
prefix	zh.wikipedia.org/zh-hk/\$term	36	
prefix	zh.wikipedia.org/zh-sg/\$term	26	
prefix	zh.wikipedia.org/zh-hans/\$term	29	
prefix	zh.wikipedia.org/zh-hant/\$term	27	
prefix	en.wikipedia.org/wiki/\$term	20	
<i>prefix corner cases:</i>			
prefix	wiki/\$term	1	20th_anniversary_Tiananmen_square_incident_march
prefix	pedia.org/zh-hant/\$term	1	User:Liangent-bot/Base64URL/5YWt5Zub5LqL5Lu2
prefix	zh-cn/\$term	1	User:Liangent-bot/Base64URL/5YWt5Zub5LqL5Lu2
prefix	.wikipedia.org/wiki/\$term	1	Princeling
prefix	gan.wikipedia.org/wiki/\$term	1	盤古樂團
prefix	zh-yue.wikipedia.org/wiki/\$term	4	劉曉波, 丁子霖, 六四事件, 天安門事件
SELF (non-URL)	\$term	18	
SELF (URL)	\$term	37	

Table 4.2: GFW rulebook summary

³中華民國總統: president of Republic of China (a.k.a. Taiwan), in traditional Chinese.

⁴延安日记: *Diary in Yan'an*, a book by Soviet diplomat Peter Vladimirov, which covers the history of Chinese Communist Party from 1942 to 1945. Published in English with titles Chinese Special Zone: 1942-1945 and The Vladimirov Diaries.

4.1 Encodings

In China, there are three encoding schemes for Chinese characters commonly in use: UTF-8, GBK, Big5. UTF-8 is the international standard. GBK is for simplified Chinese and it is the predominant encoding scheme in mainland China. Big5 is for traditional Chinese and is used mainly by Taiwan and Hong Kong websites. GBK and Big5 were developed from very early on, and they use 2 bytes to represent each Chinese character, compared to 3 bytes by UTF-8, so they are still in wide use.

Wikipedia seems to only support UTF-8, so the main testing in this study is in UTF-8. However, I did study the other two encoding schemes extensively and found that GFW also has GBK/Big5 on their rulebook. *Each rule is actually three rules*, except for cases where there is no applicable GBK or Big5 encoding.

For example, `zh.wikipedia.org & 汪洋5` is on GFW rulebook, there are three string matching rules deployed to the GFW devices:

- `zh.wikipedia.org & %E6%B1%AA%E6%B4%8B` (UTF-8)
- `zh.wikipedia.org & %CD%F4%D1%F3` (GBK)
- `zh.wikipedia.org & %A8%4C%AC%76` (Big5)

4.2 The 64-byte Limit

This study also revealed that GFW’s string matching rules have a hard 64-byte limit. All rules that are longer than 64 bytes are truncated to 64 bytes at deployment. This is a supporting evidence that GFW devices are custom made hardware, built to run string matching highly efficiently.

The first case that came to my attention is the page “`zh.wikipedia.org/wiki/2013年南方周末新年献词被删改事件6`”. This string triggers GFW reset, but it turned out that the last character 件 is not required. Counting bytes show that the full string is 68 bytes, removing the last character reduces it to 65. In fact, we test the byte strings and we see that the first 64 bytes are the culprit, e.g., the 63-byte string “`zh.wikipedia.org/wiki/2013%E5%B9%B4%E5%8D%97%E6%96%B9%E5%91%A8%E6%9C%AB%E6%96%B0%E5%B9%B4%E7%8C%AE%E8%AF%8D%E8%A2%AB%E5%88%A0%E6%94%B9%E4`” does not trigger reset, but the 64-byte string “`zh.wikipedia.org/wiki/2013%E5%B9%B4%E5%8D%97%E6%96%B9%E5%91%A8%E6%9C%AB%E6%96%B0%E5%B9%B4%E7%8C%AE%E8%AF%8D%E8%A2%AB%E5%88%A0%E6%94%B9%E4%BA`” does, where the last two bytes “`%E4%BA`” are the first 2 bytes of the UTF-8 code for the character 事 (“`%E4%BA%8B`”).

Furthermore, we examine the GBK encoding of this string. GBK encoding represents each Chinese character by two bytes, so the full string is within the 64-byte limit. Indeed, the full string is required to trigger GFW reset: “`zh.wikipedia.org/wiki/2013%C4%EA%C4%CF%B7%BD%D6%DC%C4%A9%D0%C2%C4%EA%CF%D7%B4%CA%B1%BB%C9%BE%B8%C4%CA%C2%BC%FE`”. If we remove the last character or the last byte then it won’t trigger.

Combining these data points, we can see that when GFW operator receives an order, it considers all three encodings, and if any exceeds 64 bytes, it’ll truncate that one at 64 bytes.

There are actually quite a few such cases in GFW rulebook. Combining the byte probing and examination of GBK/Big5 probing, we can identify the actual order that the GFW operator received. See Table 4.3.

5 GFW’s HTTP Response Scan: Methodology and Rulebook

5.1 Methodology (HTTP Response Scan)

We use a combination of the following to study GFW’s HTTP response filtering.

- *cURL* is a computer software project for transferring data on the web. In particular, we use the command line tool *curl* to fetch web content. For this research, the most desirable feature of *curl* is the `-r` flag, which specifies a byte range of the page content for transferring. Wikipedia content is static, so repeated *curl* commands with

⁵汪洋: Wang Yang, Vice Premier of China, member of the 17th and 18th Politburo.

⁶2013年南方周末新年献词被删改事件: 2013 Southern Weekly Incident, a conflict between the Southern Weekly editors and the Party propaganda branch.

The 64-byte Limit: Triggering String and the Actual Order	
Triggering string (UTF-8)	zh.wikipedia.org/wiki/2013年南方周末新年献词被删改事
Actual order	zh.wikipedia.org/wiki/2013年南方周末新年献词被删改事件
Triggering string (UTF-8)	zh.wikipedia.org/zh-tw/2013年《南方周末》新年特刊被删
Actual order	zh.wikipedia.org/zh-tw/2013年《南方周末》新年特刊被删改事件
Triggering string (UTF-8)	zh.wikipedia.org/zh-hans/北京中民信房地产开发有限公司
Actual order	zh.wikipedia.org/zh-hans/北京中民信房地产开发有限公司
Triggering string (UTF-8)	zh.wikipedia.org/wiki/Wikipedia:删除投票和请求/2007年6月
Actual order	zh.wikipedia.org/wiki/Wikipedia:删除投票和请求/2007年6月4日
Triggering string (UTF-8)	zh.wikipedia.org/wiki/Wikipedia:当前的破坏/存档/2010年7-
Actual order	zh.wikipedia.org/wiki/Wikipedia:当前的破坏/存档/2010年7-12月
Triggering string (UTF-8)	zh.wikipedia.org/wiki/Talk:2013年《南方周末》新年献词
Actual order	zh.wikipedia.org/wiki/Talk:2013年《南方周末》新年献词被删改事件
Triggering string (UTF-8)	zh.wikipedia.org/zh-hant/Talk:2013年《南方周末》新年献
Actual order	zh.wikipedia.org/zh-hant/Talk:2013年《南方周末》新年献词被删改事件
Triggering string (UTF-8)	zh.wikipedia.org/wiki/Talk:2013年《南方周末》新年特刊
Actual order	zh.wikipedia.org/wiki/Talk:2013年《南方周末》新年特刊被删改事件

Table 4.3: Long strings that are truncated due to the 64-byte limit.

refined byte ranges can pinpoint the exact location of the page content that offends GFW. Note that for each curl test, we need to run a “follow up” innocent curl request to examine if the connection has been reset or not.

- Custom web server. We set up a simple Apache web server so that we can freely tweak the content for transferring. This is indispensable to confirm the exact string matching rules.
- Editing Wikipedia content. This is necessary because we found that a subset of GFW’s HTTP response filtering rules only affect the Wikipedia site, not my custom web server, or other websites. Since we do not want to pollute Wikipedia content, this is done in user testing pages.
- HTTP proxy. Probing via a proxy in China can give us some information about user experience if the HTTP request and response are between that proxy IP and a target server IP. Using multiple proxies can also increase our probing efficiency. A web search gives many public proxies, for example, the list at <http://spys.ru/free-proxy-list/cn/>. These proxies can help to reveal topology of GFW operations. We found that different proxies may give different results when fetching offending content from non-offending URLs, which suggests that GFW’s HTTP response filtering is more distributed, and likely deployed at multiple locations at ISP level. On the contrary, for HTTP request filtering, probing offending or non-offending URLs, all give the same results for the many different proxies we tested (allowing for a small amount of noise in testing). This suggests that GFW’s HTTP request filtering most likely happens at the border in a controlled and integrated manner.

Regarding the candidate “offending” content for probing, we use two sources.

- Greatfire.org’s list of blocked Wikipedia pages. Based on my study of GFW’s rulebook for HTTP request filtering, we know the URLs for some of those unaccessible pages do not trigger GFW reset, so we suspect that the page content is the culprit. We run curl for these pages, and in an iterative fashion, locate the exact byte positions in the content that trigger GFW reset. Appendix A is the summary of the diagnosis of Greatfire.org’s Wikipedia list. The full list and diagnosis is available at <http://goo.gl/jUJpHb>.
- Our previous study of GFW’s HTTP request filtering gives an extensive list of sensitive terms. We put these terms, together with many derived terms in pages on our custom web server, and also on a Wikipedia test page. Then we test these pages (from many proxy IPs) to see if they trigger GFW reset. We also throw in various HTML content, for example, those of search engine result pages, to our testing pages.

Note that a page may contain multiple phrases that can trigger GFW, also some rules require two parts (i.e. “\$term_A & \$term_B”) or even three. A lot of care is needed to make precise and conclusive claims. Adding to the difficulty is that different (proxy) IPs may be subjecting to different rulebook and the resetting rates vary a lot among these (proxy) IPs and they are all much lower than 100%. The testing effort is significantly bigger than the study of GFW's HTTP request filtering, per rule. Luckily, the response filtering rulebook is two orders of magnitude smaller, so we are able to get an accurate picture of this messy operation.

5.2 GFW's HTTP Response Filtering Rulebook

In our study, we are able to identify 19 rules that GFW applies to HTTP responses. See Table 5.1. These terms can be seen as the ultimate sensitive topics China authority wants to block. Note that these rules are fully vetted, i.e., substrings or variations not listed here will not trigger GFW. Also, even though we can not claim this list is complete, there should not be any obvious omissions, because we tested numerous “sensitive” content. For example, there is no Tiananmen Incident-related term here. We have examined many pages (on or off Wikipedia site) on this topic and confirmed it.

A: Rules targeting Wikipedia only	
falun *	as in Falun Gong, the most well known GFW term
flg *	acronym for Falun Gong
法轮功	Falungong
达赖喇嘛	Dalai Lama
自由西藏	Free Tibet
西藏独立	Tibetan independence
西藏流亡政府	Tibetan government in exile
雪山狮子旗	snow mountain and lion flag
B: Rules partially deployed	
法轮大法 & 明慧网	Falun Dafa & Minghui web
新唐人电视台 & 神韵艺术团	New Tang Dynasty TV station & Shen Yun performing arts group
第二阶段 & 中央警卫局*	second stage & CCP Central Guard Bureau
胡海峰 & 威视 *	Hu Haifeng & Nuctech (in traditional Chinese)
C: Rules with (near-)complete deployment	
纳米比亚 & 威视	Namibia & Nuctech (in simplified Chinese)
纳米比亚 & 胡海峰	Namibia & Hu Haifeng
namibia & nuctech	
namibia & huhai Feng	
dongtaiwang.com	a website hosting Freegate
阅后即焚 & gfw *	burn after reading & gfw
李长春 & google.com & images/nav_logo *	

Table 5.1: GFW rulebook for HTTP response filtering. Terms with * are explained in more details.

We examined these terms on Wikipedia pages and on my custom web server and probed from hundreds of proxies in China. They can be grouped into the following three types.

- 8 rules only affect Wikipedia.org. 3 are about Falungong and the other 5 are about Tibet. A few highlights are:
 - “falun” is probably the most well known GFW term. There are interesting anecdotes about the Sweden city Falun. There were evidences that this string was applied to all internet traffic by GFW in the past, but currently non-Wikipedian pages that contain this string seem to no longer trigger GFW reset. I tested this on my custom web server, and also probing various pages about the Sweden city Falun from a large number of proxy IPs in China.

- “flg” is the acronym of Falungong. This string is very short so it causes numerous unintended damages. For example, many of those seemingly innocent Wikipedia articles reported as blocked by Greatfire.org are due to this tiny string, including articles for “Graphene”, “Same-sex_marriage”, “homosexuality”, “Human_rights”, “Comparison_of_tablet_computers”, “莫言”⁷, “日本地理”⁸, “冷战”⁹ and many more. This is particularly an issue because Wikipedia articles usually contain web links which contain the forward slash sign (“/”) which is escaped as “%2F”, e.g., the link www.barackobama.com/pdf/lgbt.pdf is www.barackobama.com%2Fpdf%2Flgbt.pdf in the HTML source for the article “Same-sex_marriage”, making the page inaccessible from China.
- The term “达赖喇嘛” (Dalai Lama) shows up in the template {{藏传佛教}} (Tibetan Buddhism), and this template exists on many Tibet-related articles. The content of this template is hidden by default, but the term exists in the page’s HTML source. This effectively makes the majority of Tibet-related pages inaccessible in China, even though most of these pages have nothing to do with Dalai Lama.
- 4 rules which are partially deployed. They trigger GFW reset only from a subset of proxy IPs, notably from the CERNET (China Education and Research Network) but not limited to CERNET. Note that if these terms are on Wikipedia.org, they trigger GFW reset from all proxy IPs I tested. Two terms in this category are about Falungong, the other two are noteworthy:
 - “胡海峰 & 威视”. The first part is Hu Haifeng, son of Hu Jintao and the second part is a company Nuctech in traditional Chinese. This is due to a scandal regarding alleged bribery by Nuctech to Namibia officials. There are four rules in the “(near-)complete deployment” category about this scandal, two in English and two in simplified Chinese. I examined this thoroughly, other combinations of terms and traditional/simplified versions are fine. My hypothesis is that the Chinese censors did study web content carefully. First they found that web pages without “纳米比亚” or “Namibia” are “clean”, so just deployed the four rules, but later they discovered that some pages with “胡海峰” and “威视” in the traditional form are also “bad”, so this rule was deployed separately, and deployed to a different set of hardware.
 - “第二阶段 & 中央警卫局”. The first part is “second stage” and the second part is “CCP Central Guard Bureau”. This is a very odd case but it is fully vetted. This rule was discovered from examining the Wikipedia article <http://zh.wikipedia.org/wiki/中共> (“中共” is the short name for Chinese Communist Party) which is reported by Greatfire.org as blocked. There is web content on the internet about somebody who claims to be from CCP Central Guard Bureau and leaks “secrets” of top Party officials, but I do not know whether this rule was motivated by that content or this Wikipedia article. This arcane case suggests that there might be more of such odd cases that my current investigation did not catch.
- 7 rules which are deployed to cover all Chinese users. For almost all the proxy IPs I tested, these terms trigger GFW reset. Four of these are about the aforementioned Hu Haifeng-Nuctech-Namibia scandal, one is a website which hosts the GFW-circumvention tool Freegate, the other two are noteworthy:
 - “阅后即焚 & gfw”. The first part is “burn after reading”. This rule was discovered from examining the Wikipedia article <http://zh.wikipedia.org/wiki/中国网络软件过滤关键字列表>¹⁰. There is a widely circulated article online with the title “阅后即焚- GFW的来世今生” (“Burn after reading - the past life and present form of GFW”) since 2009. The author is anonymous but the article contains unprecedented details of GFW operations. This rule shows the extreme sensitivity of this article.
 - “李长春 & google.com & images/nav_logo”. The first part is Li Changchun, China’s Propaganda Chief from 2002 to 2012. This rule was discovered when we put Google search result page’s content together with all the known sensitive terms in a probing target. This rule is significant in that it is specifically devised to target Google search pages which contain the name “Li Changchun”. The third part “images/nav_logo” is put in place to ensure it is a Google search result page. Li Changchun is the only Party officials that receives this “treatment”.

⁷莫言: Mo Yan, Chinese author who won the Nobel Prize in Literature in 2012.

⁸日本地理: Japanese geography.

⁹冷战: Cold War.

¹⁰中国网络软件过滤关键字列表: List of keywords filtered by China’s internet censorship software.

5.3 Learnings and Mysteries of GFW's HTTP Response Filtering

Unlike the HTTP request filtering, it is very hard to study GFW's exact mechanism to scan HTTP responses, because it is fairly straightforward to construct various HTTP requests for probing, but much harder to construct HTTP responses at will. In particular, I have not been able to find a way to forge the source IP address in the IP packet header. Therefore, even though we have obtained a quite comprehensive list of filtering rules for GFW's HTTP response scan in this study, we have obtained only some basic understanding on GFW's practice of HTTP response scanning.

- GFW's HTTP request filtering and response filtering are two separate systems. For one, their filtering rules are entirely different. For two, GFW's HTTP request filtering is homogeneous and has near perfect trigger rate, but GFW's HTTP response filtering varies hugely, not only in the triggering rates, but also in the filtering rules in effect. For example, CERNET (Chinese Education and Research Network) seems to have all the rules in place, but some other ISPs only have a subset. The heterogeneity in filtering rules, i.e. different ISPs are subject to different filtering rules, actually provides an opportunity to study GFW topology. This could be a future research topic.
- One remarkable finding is that GFW does not just look at individual TCP packet, but instead, it "remembers" the entire TCP session to look for offenders. This becomes evident when the filtering rule is "\$term_A & \$term_B", and the two terms show up far apart (hundreds of thousands bytes from each other) on a webpage, GFW will still be able to reset the connection. To achieve this requires significant investment in infrastructure, and it is probably also the reason why the rulebook is so much smaller for HTTP response filtering than HTTP request filtering.
- Given the list, I also studied the GBK and Big5 encodings for each of the rules. 13 terms have proper GBK encoding and 1 term has proper Big5 encoding. Indeed, when the UTF-8 version triggers GFW reset, the corresponding GBK or Big5 version will also. This is consistent with GFW's HTTP request scan, in which GBK and Big5 encodings are filtered as well.
- We studied upper cases and lower cases of these rules. GFW's HTTP response filtering is case insensitive, which is consistent with GFW's HTTP request filtering.

Despite these learnings, we have a major unknown issue and a few puzzling mysteries.

- We don't know exactly how GFW specifically targets Wikipedia site for these rules. I tried putting those Type A terms in a page on the custom web server but that does not trigger GFW reset. I also tried adding the strings "zh.wikipedia.org", "en.wikipedia.org", and Wikipedia.org IP addresses (in digits and in byte format) to the page, still no resets. My hypothesis is that GFW looks at the source IP address field in the IP packet header to look for the Wikipedia.org IP, but we are unable to confirm this. Interested readers and researchers are welcome to study this more deeply.
- A great mystery is the "masking effect" of the term "法轮" (falun). The phenomenon is the following, we fetch a (non-Wikipedia) page which contains an offending text, it will trigger GFW reset as expected, however, if we add the term "法轮" to that page, fetching the page will no longer trigger GFW reset. It's also the case when we specify a byte range with curl, i.e. if that byte range contains an offending string and the term "法轮", it will not trigger GFW reset. This phenomenon is reproducible for all the rules (except for the rule "法轮大法 & 明慧网" which contains "法轮" itself) on a non-Wikipedia page from many (but not all) proxy IPs we tested. It is as if that the existence of "法轮" makes the GFW filtering device go into a different mode. My speculation is that GFW may go into a surveillance mode when it sees "法轮", hence skipping the resetting, either intentionally or due to a bug. Interested readers and researchers are welcome to study this more deeply.
- One other mystery is about the term "flg" on Wikipedia.org. This string on Wikipedia.org site does trigger GFW reset. However, if we run curl to only fetch these three bytes (by the "-r" flag), it will not trigger GFW reset. If we include something extra by making the byte range more than 3, then it will trigger GFW reset. This extra thing can be anything - some arbitrary characters or whitespace or punctuations. Other longer rules (e.g. the five-bytes rule "falun") do not have this property. I do not know the reason for this phenomenon. It could be a result of GFW special mechanism for string matching.

6 Wikipedia Specifics

6.1 Wikipedia Content Structure

When people talk about Wikipedia, they often just mean the *articles* (or, “real” content). These are the pages a normal Wikipedia visitor reads. But underneath that, Wikipedia actually hosts a lot more types of content. These types are called *namespace*. Table 6.1 has more details, based on information from <http://www.mediawiki.org/wiki/Manual:Namespace>.

Index & Name	Purpose	Counts (zh en)	URL
0: (Main)	“Real” content; articles	1,177,122	[en zh].wikipedia.org/wiki/\$page_title
1: (Main) talk		172,707 5,215,221	[en zh].wikipedia.org/wiki/Talk:\$page_title
2: User	User pages	58,497	[en zh].wikipedia.org/wiki/User:\$page_title
3: User talk		562,241 9,061,249	[en zh].wikipedia.org/wiki/User_talk:\$page_title
4: Project	Information about the wiki	27,953	[en zh].wikipedia.org/wiki/Wikipedia:\$page_title
5: Project talk		4,097 210,611	[en zh].wikipedia.org/wiki/Wikipedia_talk:\$page_title
6: File	Media description pages	35,583	[en zh].wikipedia.org/wiki/File:\$page_title
7: File talk		4,162 176,685	[en zh].wikipedia.org/wiki/File_talk:\$page_title
8: MediaWiki	Site interface customization	6,846	www.mediawiki.org/wiki/MediaWiki:\$page_title
9: Mediawiki talk		299 1,081	www.mediawiki.org/wiki/MediaWiki_talk:\$page_title
10: Template	Template pages	861,514	[en zh].wikipedia.org/wiki/Template:\$page_title
11: Template talk		6,226 204,302	[en zh].wikipedia.org/wiki/Template_talk:\$page_title
12: Help	Help pages	303	[en zh].wikipedia.org/wiki/Help:\$page_title
13: Help talk		66 628	[en zh].wikipedia.org/wiki/Help_talk:\$page_title
14: Category	Category description pages	153,155	[en zh].wikipedia.org/wiki/Category:\$page_title
15: Category talk		14,739 677,383	[en zh].wikipedia.org/wiki/Category_talk:\$page_title
others	custom namespaces	6,946 160,434	

Table 6.1: Wikipedia namespace; counts are based on the 2013-09-08 dump (zh) and the 2013-09-04 dump (en).

In the beginning, I was considering only to examine the articles (i.e. namespace=0). Then an interesting case came to my attention. I saw that Greatfire.org’s Wikipedia tracking page shows that the page <http://zh.wikipedia.org/wiki/Wikipedia:頁面存廢討論>¹¹ is blocked. I verified that it is indeed the case. This page is in the `Project` namespace, so I expanded my examination from Articles only to all applicable namespaces. To be precise, for the Chinese version, we tested all entries except for MediaWiki (namespace=8), MediaWiki talk (namespace=9) and those “customized” namespaces (namespace >=100). For the English version, we tested entries in Article (namespace=0), User (namespace=2), Project (namespace=4), File (namespace=6), Template (namespace=10), Help (namespace=12), Category (namespace=14), i.e., we ignored those “Talk” namespaces for English.

Indeed, we found 7 GFW rules against `Project`, 25 rules against `User`, 2 rules against `Category`, 2 rules against `File`, and 23 rules against `Talk`.

Also, Wikipedia has smart localization schemes for these namespaces. For example, to access the `User` page for the Chinese version, one can use the English URL: “zh.wikipedia.org/wiki/User:\$user”, or the localized URL: “zh.wikipedia.org/wiki/用戶:\$user”. The latter actually redirects to the former (using HTTP 301 redirect). I examined the localized version of the 25 rules against `User`, none of them are on GFW rulebook, so I decided to ignore these localized URLs.

6.2 Nine Variants of the Chinese Wikipedia

Usually we think Chinese Wikipedia as “one version”, but actually Wikipedia has two “main variants” and five “fine variants” of Chinese (Table 6.2). The two main variants are simplified Chinese and traditional Chinese. These two

¹¹ 頁面存廢討論: discussion for articles deletion. Its English counterpart is en.wikipedia.org/wiki/Wikipedia:Articles_for_deletion

are deprecated, but their legacy URLs still work. The five current finer variants are: 大陆简体 (Mainland Simplified), 台灣正體 (Taiwan Traditional), 香港繁體 (Hong Kong Traditional), 澳門繁體 (Macau Traditional), 马新简体 (Malaysia/Singapore Simplified). Currently Wikipedia uses the canonical URL: `zh.wikipedia.org/wiki/$term` and the rendering follows user's preference setting, and the user can manually select a version from a pull-down menu, which overrides the default setting, and with a distinctive URL like `zh.wikipedia.org/zh-cn/$term`. Furthermore, there is a general URL like `zh.wikipedia.org/zh/$term` which renders just like the canonical URL `zh.wikipedia.org/wiki/$term`. So the total number of valid URL patterns that we need to test is 9.

Variant	URL	GFW Rule Counts
Canonical	<code>zh.wikipedia.org/wiki/\$term</code>	259
Equivalent alias	<code>zh.wikipedia.org/zh/\$term</code>	35
大陆简体 (Mainland Simplified)	<code>zh.wikipedia.org/zh-cn/\$term</code>	85
台灣正體 (Taiwan Traditional)	<code>zh.wikipedia.org/zh-tw/\$term</code>	51
香港繁體 (Hong Kong Traditional)	<code>zh.wikipedia.org/zh-hk/\$term</code>	36
澳門繁體 (Macau Traditional)	<code>zh.wikipedia.org/zh-mo/\$term</code>	0
马新简体 (Malaysia/Singapore Simplified)	<code>zh.wikipedia.org/zh-sg/\$term</code>	26
简体 (Simplified)	<code>zh.wikipedia.org/zh-hans/\$term</code>	29
正體 繁體 (Traditional)	<code>zh.wikipedia.org/zh-hant/\$term</code>	27

Table 6.2: Variants of Chinese Wikipedia. Rule counts exclude those few corners cases.

To the GFW devices, these are all different URLs, so a rule like `zh.wikipedia.org/wiki/陈光诚12` will not be able to block a user request like `zh.wikipedia.org/zh-cn/陈光诚`. However, the broad rule like `zh.wikipedia.org & 刘晓波13` can block all these variants since their URLs all contain the string `zh.wikipedia.org`. GFW uses broad rules for many terms, but it does use prefix rules for many other terms, covering only a subset of these variants. This may seem puzzling to an outside observer, but there are two possible explanations for this phenomenon. Firstly, the current “incomplete” block does its job well, because majority of Wikipedia visits are using the canonical path, and GFW is never assumed to be watertight. Secondly, it is possible that GFW operators simply accept “orders” from their supervising authorities, i.e. the censorship officials. These officials may simply order: “block 刘晓波 on Wikipedia”, for which case the GFW operators will use the broad rule; or, they may pass on one URL or a list of URLs for GFW to block, e.g. `zh.wikipedia.org/zh-cn/陈光诚`, in which case the GFW operators would just deploy this string.

We tested all 9 variants thoroughly in this study. See Table 6.2 for the counts for rules targeting each variant. There are 259 rules for the canonical path. For the eight non-canonical variants, there are 289 rules, covering 161 unique terms. It is noteworthy that 78 out of these 161 terms only have one or more non-canonical variants blocked, but not the canonical one. In my view, this is a strong indication that GFW operators and/or their supervising authority have poor understand of Wikipedia.

A little more history about these variants. According to http://en.wikipedia.org/wiki/Chinese_wikipedia, in the very beginning, the Simplified and Traditional Chinese Wikipedias are entirely separate. But the content between the two is inevitably similar, so in 2005, it was decided to merge the two and create a unified Chinese Wikipedia. This effort is quite a success. It handles the simplified-traditional conversion seamlessly, and also handles different vocabularies among the five regions.

6.3 Wikipedia's Handling of Traditional and Simplified Chinese Characters

The Chinese language has mainly two writing systems: Traditional Chinese and Simplified Chinese. It is usually a 1-to-1 correspondence for the common characters, but there are many cases that multiple traditional Chinese characters are mapped to the same simplified character.

¹²陈光诚: Chen Guangcheng, the legendary blind civil rights activist, whose escape from house arrest in Apr 2012 raised international attention.

¹³刘晓波: Liu Xiaobo, the 2010 Nobel Peace Prize laureate.

In the digital world, a traditional character and its simplified counterpart are just two different objects, and represented by entirely different encodings. For example, “漢”, the traditional character’s UTF-8 encoding is %E6%BC%A2, and its simplified version “汉”, has UTF-8 code of %E6%B1%89.

For GFW, blocking one version does not block the other version. In fact, we see numerous cases where one is blocked while the other is not. For a target string, GFW operators have to take the effort to find and deploy both versions in order to block both the traditional and simplified versions.

In the Wikipedia land, the Wikipedians take a smart and ambitious approach - automatic traditional-simplified conversion. This is a rather nontrivial task, because there are a lot of nitty gritty details in the conversion, luckily Wikipedia has enough volunteers and the conversion system is mostly correct and seamless, despite a small amount of inevitable complexities and confusion, which are inevitable. The page zh.wikipedia.org/wiki/Wikipedia:繁简处理¹⁴ has more details.

This automatic conversion has implications to GFW blockage. We cover this aspect in Section 6.4.

6.4 Wikipedia's Two Types of Redirects

There are two types of redirects in Wikipedia: the *soft* redirect and the *hard* redirect. The former handles the redirection behind the scene without further request from the browser. The latter involves HTTP 301 response code (301 Moved Permanently), which directs the browser to send another HTTP request for the new target.

This has implication regarding GFW blockage. Let’s consider the case where a page (A) redirects to another page (B), page A is not on GFW blacklist but page B is. If this redirect is a soft one, then the user can view page A without interruption, but if this redirect is a hard one, then the user’s connection to Wikipedia will be reset.

Soft redirects are set up in the article source using the #REDIRECT keyword. For example, the source of the English article “Tiananmen.Square.Massacre” is simply #REDIRECT[[Tiananmen Square protests of 1989]], so a page request for <http://en.wikipedia.org/wiki/Tiananmen.Square.Massacre> would render the content of the page http://en.wikipedia.org/wiki/Tiananmen_Square_protests_of_1989. This redirection happens on the server side.

Soft redirects are also used for many pairs of page titles in traditional and simplified Chinese. Prior to the traditional-simplified unification, for many subjects, two articles were created, one with title in traditional Chinese and one with title in simplified Chinese. They could have entirely separate content, or one could be a redirect to the other. During the unification effort, all these cases were merged to the same content with one article redirecting to the other. For example, Chinese Wikipedia currently has both <http://zh.wikipedia.org/wiki/北京大学>¹⁵ (simplified Chinese) and <http://zh.wikipedia.org/wiki/北京大學>¹⁶ (traditional Chinese), and the latter soft-redirects to the former.

However, for many newer pages, e.g. those created after the unification, only one form exists in Wikipedia, a page request to the other form is handled by a hard redirect (i.e. HTTP 301 redirect). For example, Chinese Wikipedia has an article for 许志永¹⁷ (simplified Chinese), but it does not have an article for the traditional form 許志永¹⁸. A page request to <http://zh.wikipedia.org/wiki/許志永> is hard redirected to <http://zh.wikipedia.org/wiki/许志永>.

Hard redirects are also used for trivial typological conversions. For example, <http://en.wikipedia.org/wiki/Crown%20Prince%20Party> (%20 is whitespace) is hard redirected to http://en.wikipedia.org/wiki/Crown_Prince_Party, (for what it is worth, the latter page is actually further soft-redirected to <http://en.wikipedia.org/wiki/Princelings>). Nonstandard upper/lower casing in system keywords also involves hard redirect, for example, <http://en.wikipedia.org/wiki/TALK:Princelings> is hard redirected to the standard form <http://en.wikipedia.org/wiki/Talk:Princelings>. Also note that page titles are case sensitive except for the first character. A nonstandard case for the first character gives a hard redirect, e.g. <http://en.wikipedia.org/wiki/princelings> is hard redirected to <http://en.wikipedia.org/wiki/Princelings>, while <http://en.wikipedia.org/wiki/PRincelings> will give a “Wikipedia does not have an article with this exact name” page.

To investigate which redirect is happening is straightforward. For soft redirect, one can see a small subtitle like (Redirected from) (English) or (重定向自) (simplified Chinese) under the title line of the main article. For hard redirect, one can open the developer tools console in the browser and check the requests the browser has made. If we see the original page request gets a status code of 301, and a new page request is made, then that tells a hard redirect has happened.

¹⁴繁简处理: Handling of traditional and simplified Chinese.

¹⁵北京大学: Peking University, in simplified Chinese.

¹⁶北京大學: Peking University, in traditional Chinese.

¹⁷许志永: Xu Zhiyong, in simplified Chinese.

¹⁸許志永: Xu Zhiyong, in traditional Chinese.

6.5 Other Chinese-related Wikipedias

Our focus in this study is the Chinese and English versions of Wikipedia, but actually there are 7 more Wikipedias that are related to Chinese, these are, roughly, “dialects” of the Chinese language. See Table 6.3.

These Wikipedias are much smaller and less popular than the Chinese one, so I did not test them exhaustively. But still, I tested all the GFW terms that affect Chinese and English Wikipedias on these versions. Indeed, we found that GFW does target the Cantonese and Gan Wikipedias for a few terms, but not the other versions.

The four terms for which the Cantonese Wikipedia is targeted are: 丁子霖¹⁹, 劉曉波²⁰, 六四事件²¹, and 天安門事件²². The one term for which the Gan Wikipedia is targeted is 盤古樂團²³. The connection here is that the Pangu band was originated from the Jiangxi Province, where the Gan dialect is dominant.

Versions	URL	GFW Rule Counts
Classical Chinese Wikipedia	zh-classical.wikipedia.org/wiki/\$term	0
Minnan Wikipedia	zh-min-nan.wikipedia.org/wiki/\$term	0
Cantonese Wikipedia	zh-yue.wikipedia.org/wiki/\$term	4
Mindong Wikipedia	cdo.wikipedia.org/wiki/\$term	0
Wu Wikipedia	wuu.wikipedia.org/wiki/\$term	0
Hakka Wikipedia	hak.wikipedia.org/wiki/\$term	0
Gan Wikipedia	gan.wikipedia.org/wiki/\$term	1

Table 6.3: 7 Wikipedias related to Chinese

7 Caveats

The author strongly urges the reader to read this section carefully, in order to get the most accurate understanding of our findings.

7.1 On Completeness

Even though the author uses the word “complete” in the title of this report, the intention here is that our list covers *all GFW string matching rules that affect the current Chinese and English Wikipedias in its HTTP request filtering*, this is the right way to interpret the word “complete”. I believe that not all terms on the GFW rulebook **targeting** the Chinese/English Wikipedias are identified, for the following scenarios:

- GFW may contain Wikipedia-related rules that do not affect any actual Wikipedia pages. This could be that the page never existed, or the page has been deleted. Since in this study we are only examining pages in the current snapshot (2013-09-08), we may miss those non-existent or deleted pages.
- GFW’s blacklist is a mess. It is highly likely that certain rules are *redundant*, i.e. one rule may be strictly dominated by another rule. For example, we know “.wikipedia.org/Princeling” is one rule, if GFW also has “en.wikipedia.org/Princeling” or “.wikipedia.org/Princelings” on its rulebook, there is no way for us to verify that. What we report in this study is fully vetted, and should cover all *verifiable* rules.
- As we have mentioned in Section 4, GFW rulebook contains many URLs which may affect Wikipedia. Only those URL terms which affect pages in the Chinese Wikipedia or articles (i.e. namespace==1) in the English Wikipedia are included in this report. These URL terms are covered in Table 8.11.

¹⁹丁子霖: Ding Zilin, organizer of the Tiananmen Mothers.

²⁰劉曉波: Liu Xiaobo, in traditional Chinese.

²¹六四事件: June 4th Incident, a.k.a. the Tiananmen Protest in 1989.

²²天安門事件: Tiananmen Incident, in traditional Chinese

²³盤古樂團: Pangu, named as “Punk God”, a Chinese avant-garde punk band banned in China, in traditional Chinese.

One more related note, our list does include certain rules that do not affect actual Wikipedia titles. This is because after we have identified a term, we will also investigate its various forms, e.g. its traditional/simplified counterpart, other variants and versions of Wikipedia. Probing these other forms may surface more GFW rules that target Wikipedia. But the pages that these rules target may not actually exist. Some of them are (hard) redirected to an actual Wikipedia page, some may give the “Wikipedia does not have an article with this exact name” page. As long as GFW has such a rule on its rulebook, it is included in this report.

In version 1.0, we did not study GFW’s HTTP response filtering, so we cautioned readers that our rulebook did not capture all Wikipedia pages that are not accessible from China. In this version, we studied GFW’s HTTP response filtering extensively and identified a small but impactful list of filtering rules. If a Wikipedia page’s HTML source offends any of these rules, China users will not be able to access that page successfully. We are reasonably sure that we have captured all GFW’s HTTP response filtering rules that affect Wikipedia, but we skipped compiling the list of affected pages. This is because 1, Wikipedia content is in flux; 2, to examine the HTML source of all Wikipedia pages is a daunting or infeasible task; and 3, unlike the HTTP request filtering list, the list of affected pages due to HTTP request filtering do not give much more information about China censor’s motivation or priorities.

If the reader encounters a blocked Wikipedia page, he can first check the complete rulebook in Section 8. If the page’s URL does not offend any rule there, he can try to get the HTML source and search for the rules in Table 5.1. Also keep in mind of the following potential issues:

- A page whose URL does not offend GFW rulebook might be inaccessible if it is “hard” redirected (see Section 6.4) to an offending page.
- Due to network topology, different ISPs/regions may connect to Wikipedia by different paths, some of which might not be affected by GFW interference.
- GFW does not operate at 100%. Certain level of randomness does exist.

Anyway, if a Wikipedia page is inaccessible, its URL does not offend the rulebook in Section ?? and its HTML source does not offend any rules in Table 5.1, the author welcomes such data points (by email) and will investigate.

7.2 Possible Reaction by GFW

This report will likely embarrass some GFW operators (if they ever see it), not only in revealing these supposed “secrets”, but also in revealing the porousness and lack of maintenance of the GFW rulebook. In a way, this study is an “audit” on GFW operators’ work. They may fix some cases, or even give it an overhaul.

In particular, I’d expect GFW to revise all those terms who are blocked on a variant path other than the canonical path. It is apparent that China censors *intended* to block these terms but the current implementation does not actually do the job. There are 78 such cases, including the high profile case of 许志永²⁴, for whom the variant zh.wikipedia.org/zh-cn/许志永 is blocked, but the standard path zh.wikipedia.org/wiki/许志永 is easily accessible.

The author will keep an eye on such possible changes, and the readers are welcome too.

²⁴许志永: Xu Zhiyong, a prominent civil rights activist. His arrest on July 16, 2013 attracted wide attention.

8 GFW Rulebook for Wikipedia

The GFW Rulebook for Wikipedia has 919 entries (excluding URL terms). I group them into the following eleven categories:

- **Events.** These are articles covering recent events. From here we can gain insights on what is deemed sensitive by China's censors. It is interesting to see that lots of events in 2010 and 2011 are blocked. The counts for 2012 and 2013 are much smaller. This suggests that China's censors were more keen on Wikipedia in 2010 and 2011, and their attention declined since.
- **Media, Publications, Censorship and Circumvention.**
- **Organizations, Political & Government.**
- **People.** I further group people terms into three subcategories: **Dissidents, Government Officials, Miscellaneous.** Note that people related to Hong Kong, Taiwan, Tibet, Uyghur and Falungong are described in respective categories. Five more people names are in the SELF category (see below).
- **Regional.** This group contains four subgroups for topics related to **Hong Kong, Taiwan, Tibet** and **Uyghur**.
- **Falungong.** This group is for terms related to Falungong, a religious group persecuted in mainland China. It is somewhat unexpected that there are only 19 Falungong-related rules, and many Falungong related articles are not on GFW's rulebook. This does not mean those articles are accessible in China though, because GFW has five filtering rules in its HTTP response scan that target Falungong.
- **Tiananmen.** This group is for terms related to the 1989 Tiananmen Protest (a.k.a., the June 4th Incident). We can see that Tiananmen Incident is the *most* sensitive topic the China censors want to block. 116 rules are in this section, including all kinds of arcane terms. There are also about 10 rules in the Government category and about 15 rules targeting the User namespace which are motivated by the Tiananmen Incident.
- **Miscellaneous.** This small group covers terms which do not naturally fall in any other categories.
- **Non-Articles.** This group contains GFW rules that target non-article pages (i.e. namespace != 0, e.g. Talk pages, User pages, etc). I guess few people would have expected that China censors tried this hard to precisely block such pages. For example, for certain topics, GFW blocks the Talk page but not the main article. Looking at the page content, we see that the Talk page contains unwelcome content while the main article does not. This suggests that China censors had examined the Wikipedia page diligently. Certain pages in this group are extremely obscure. Their presence demonstrated that China censors have intimate knowledge of Wikipedia content, or that certain Wikipedia editors informed the China censors about the existence of such content. The author is inclined to believe in the latter.
- **SELF.** This is grouping by match type, not really a grouping on subject. I separate this category because they are significant to a degree. In theory one would assume that terms blocked by this fashion must be hyper-sensitive, but this is not always the case. Quite several terms in this group are not that sensitive at all.
- **URLs.** GFW has a large number (estimated to be in the thousands or tens of thousands) of URLs on their rulebook. Here we include those URL terms that 1. affect pages in the Chinese Wikipedia, or 2. affect articles in the English Wikipedia. There are 37 of them, and not included in the total count of 919.

Obviously, many terms fall into multiple categories. For simplicity we assign one category for each term. The guideline I use here is "significance" and/or specificity. For example, if a term is both in SELF and in People-Official, I will put it in SELF because the SELF category is more significant. The rough order of significance I use is: SELF > Non-Articles > Regional > People > others.

For rules targeting the English Wikipedia, I opted to categorize them by topic, (usually) together with their corresponding Chinese entries. There are only 22 rules targeting the English Wikipedia, so I'll also list them below:

GFW Rule	Category	Note
en.wikipedia.org/wiki/Jasmine_Revolution_in_China	Events	
en.wikipedia.org/wiki/Charter_08	Events	
en.wikipedia.org/wiki/Crown_Prince_Party	Government	
.wikipedia.org/wiki/Princeling	Government	non-standard rule
en.wikipedia.org/wiki/Chen_Guangcheng	Dissidents	
en.wikipedia.org/wiki/Liu_Xiaobo	Dissidents	
en.wikipedia.org/wiki/刘晓波	Dissidents	Liu Xiaobo, in simplified Chinese
en.wikipedia.org/wiki/劉曉波	Dissidents	Liu Xiaobo, in traditional Chinese
en.wikipedia.org/wiki/Liu_Xianbin	Dissidents	The Chinese article is not blocked
en.wikipedia.org/wiki/Dalai_Lama	Tibet	
en.wikipedia.org/wiki/Tenzin_Gyatso	Tibet	
en.wikipedia.org/wiki/Flag_of_Tibet	Tibet	
en.wikipedia.org/wiki/Tibetan_Independence_Movement	Tibet	
en.wikipedia.org/wiki/Students_for_a_Free_Tibet	Tibet	
en.wikipedia.org/wiki/Rebiya_Kadeer	Uyghur	
en.wikipedia.org/wiki/East_Turkestan_Independence_Movement	Uyghur	
en.wikipedia.org/wiki/Tiananmen_Square_Protests_of_1989	Tiananmen	
wiki/20th_Anniversary_Tiananmen_Square_Incident_March	Tiananmen	non-standard rule
en.wikipedia.org/wiki/Tank_Man	Tiananmen	
en.wikipedia.org/wiki/Tiananmen_Massacre	Tiananmen	
en.wikipedia.org/wiki/Tiananmen_Papers	Tiananmen	
en.wikipedia.org/wiki/Great_Wall_of_China	Miscellaneous	

Table 7.0: GFW Rulebook: Rules Targeting the English Wikipedia

The following is the table format and notation for Section 8.1 to Section 8.11.

- **prefix.standard**: means the GFW rule is zh.wikipedia.org/wiki/\$term.
- **broad.standard**: means the GFW rule is zh.wikipedia.org & \$term.
- **blue texts** are page titles of (with link to) Chinese Wikipedia pages whose canonical path is blocked.
- **green texts** are page titles of (with link to) Chinese Wikipedia pages whose canonical path is not on GFW rulebook but a variant is. rulebook.
- **brickred texts** are page titles of (with link to) blocked English Wikipedia pages.
- The first number in the last column denotes the number of Chinese Wikipedia pages that offend GFW rulebook. A second optional number prefixed with “EN:” denote number of such English Wikipedia pages.
- As explained in Section 7.1, some entries do not affect any Wikipedia title, hence the count 0.
- If a term is identical to a blocked page title and the rule is on the canonical path (i.e. “/wiki/”), we color the term in the first column and do not repeat in the third column.
- When a term is targeted by multiple rules, we do not repeat the term on subsequent rows. We use the following order: “/wiki/” > “/zh/” > “/zh-cn/” > “/zh-tw/” > “/zh-hk/” > “/zh-sg/” > “/zh-hans/” > “/zh-hant/”.
- Horizontal lines separate different entries. Terms which are close in meaning are grouped together. For example, if both the simplified term and traditional term in on GFW rulebook, they are grouped together.

8.1 Events

Entries in this section are in reverse chronological order. I further divide the terms into the following buckets:

- Events in 2012-2013, 3 events, 6 rules.
- Events in 2011, 7 events, 26 rules. In particular, the Chinese Jasmine Revolution alone has 12 rules including one targeting the English Wikipedia, which speaks for its degree of sensitivity from China authority's perspective.
- Events in 2010, 6 events, 14 rules.
- Events 2005-2009, 7 events, 13 rules. In particular, Charter 08 has 6 rules, including one targeting the English Wikipedia, which speaks for its degree of sensitivity from China authority's perspective.
- Events before 2000, 6 events, 10 rules. It is both puzzling and funny that a historical event in 1085 (during the Song Dynasty) is on GFW's rulebook. That event is not well known and the page content is not sensitive either.

Term	GFW Rule	Pages Affected
Events in 2012-2013		
七不讲	prefix.standard	1
<i>Seven Do-not-talk, seven areas that China authority forbids discussion, outlined in an internal memo on May 13, 2013. They are: 1, universal values; 2, freedom of press; 3, civil society; 4, civil rights; 5, historical mistakes of CCP; 6, crony capitalism; 7, judicial independence.</i>		
南周事件	prefix.standard	1
2013年南方周末新年献词被删改事	prefix.standard	1 2013年南方周末新年献词被删改事件
2013年《南方周末》新年献词	prefix.standard	2 2013年《南方周末》新年献词被删改事件
2013年《南方周末》新年特刊被删	zh.wikipedia.org/zh-tw/\$term	1 2013年《南方周末》新年特刊被删改事件
<i>Southern Weekly Incident, January 2013, a major conflict between Southern Weekly editorial staff and the propaganda authority.</i>		
重庆万盛事件	prefix.standard	1
<i>a protest in Wansheng, Chongqing, on Apr 12, 2012.</i>		
Events in 2011		
烏坎事件	prefix.standard	1
	zh.wikipedia.org/zh-cn/\$term	
	zh.wikipedia.org/zh-tw/\$term	
2011年陸豐烏坎事件	prefix.standard	1
	zh.wikipedia.org/zh-cn/\$term	
<i>Wukan Incident, a large scale anti-corruption protest in Wukan, Guangdong from Sep 2011 to 2012.</i>		
五毛蛋	prefix.standard	3 五毛蛋争议 五毛蛋事件
<i>"Fifty-cent egg". On Oct 1, 2011, during then-Premier Wen Jiabao's visit to a university, he asked for the price of an egg and got a reply of fifty cents (RMB), which was deemed as too low by the general public.</i>		
大连反对PX项目游行	zh.wikipedia.org/zh-cn/\$term	1 大连反对PX项目游行
<i>a protest in Dalian, Liaoning Province on Aug 14, 2011 against a PX chemical project.</i>		
2011年温州动车追尾事件	prefix.standard	1
2011年杭深线动车组列车追尾事故	zh.wikipedia.org/zh-tw/\$term	1 2011年杭深线动车组列车追尾事故
<i>A high speed rail accident near Wenzhou, Zhejiang on July 23, 2011. It has long lasting effect on China's rail system and control of press.</i>		
广州市新塘事件	prefix.standard	1
<i>A riot in Xintang Township, Zengcheng, Guangdong Province on Jun 10, 2011. The trigger was a street peddler beaten by Chengguan, despite the victim being pregnant.</i>		
内蒙古抗议	prefix.standard	2 内蒙古抗议示威事件
	zh.wikipedia.org/zh-tw/\$term	

continued on next page...

...continued from previous page		
Term	GFW Rule	Pages Affected
2011年内蒙古抗议示威事件	zh.wikipedia.org/zh-hk/\$term prefix.standard	1
zh.wikipedia.org/zh-cn/\$term <i>a protest in Inner Mongolia in May 2011 against the killing of a Mongolian environmental activist.</i>		
茉莉花革命	prefix.standard	1
茉莉花事件	zh.wikipedia.org/zh-hk/\$term prefix.standard	1
茉莉花行动	prefix.standard	1
中国茉莉花革命	zh.wikipedia.org/zh-sg/\$term prefix.standard	1
中國茉莉花革命	zh.wikipedia.org/zh-cn/\$term prefix.standard	1
中國茉莉花活動	zh.wikipedia.org/zh-cn/\$term prefix.standard	1
Jasmine Revolution in China	zh.wikipedia.org/wiki/\$term	0-EN:1 Jasmine Revolution in China(en)
<i>various terms for the Jasmine Revolution (in China), a series of nationwide protests since February 2011.</i>		
Events in 2010		
2010-2011年突尼	prefix.standard	6 2010-2011年突尼斯騷亂 2010-2011年突尼斯骚乱 2010-2011年突尼西亚反政府示威 2010-2011年突尼西亚骚乱 ...
<i>Tunisia protests from December 2010 to 2011, a.k.a. Jasmine Revolution.</i>		
Wikileaks	zh.wikipedia.org/zh-hk/\$term	3 WikiLeaks WikiLeaks泄露美国外交电报事件 Wikileaks
维基解密泄露美国外交电报事件	zh.wikipedia.org/zh-hk/\$term	1 维基解密泄露美国外交电报事件
Wikileaks泄露美国外交电报事件	zh.wikipedia.org/zh-cn/\$term	1 WikiLeaks泄露美国外交电报事件
zh.wikipedia.org/zh-tw/\$term <i>The Wikileaks incident. These Chinese Wikipedia pages cover part of those cables related to China and Taiwan.</i>		
2010年中国反日示威活动	zh.wikipedia.org/zh-hk/\$term	1 2010年中国反日示威活动
zh.wikipedia.org/zh-sg/\$term <i>Anti-Japan protests in October 2010 in China.</i>		
诺贝尔和平奖	zh.wikipedia.org/zh-cn/\$term	1 诺贝尔和平奖
2010年诺贝尔和平奖	zh.wikipedia.org/zh-tw/\$term zh.wikipedia.org/zh-hk/\$term zh.wikipedia.org/zh-hant/\$term	2 2010年诺贝尔和平奖 2010年诺贝尔和平奖事件
<i>Nobel Peace Prize in 2010, which was awarded to Liu Xiaobo on Oct 8, 2010 and outraged Chinese authority.</i>		
颱風凡亞比	prefix.standard	2 颱風凡亞比_(2010年)
<i>Typhoon Fanapi. Not politically sensitive, but the page contains casualty information in mainland China.</i>		
谷歌退出中国事件	prefix.standard	1
zh.wikipedia.org/zh-cn/\$term <i>Google "exiting" China, i.e. stopping self-censorship on google.cn, in March 2010.</i>		
Events in 2005-2009		
旭日玩具厂群体斗殴事件	prefix.standard	1
<i>A violent dispute between migrant Uyghurs and Han workers at a toy factory on Jun 25/26, 2009, a.k.a. Shaoguan Incident. It was regarded as one trigger of the violent July 5th Riot in Xinjiang.</i>		
威视腐败案	prefix.standard	1
<i>The Nuctech-Namibia scandal, Nuctech is a Chinese tech company alleged to have bribed Namibia officials. The case involved Hu Haifeng, son of Hu Jintao and is heavily censored on the web.</i>		
汶川大地震死难学生调查活动	zh.wikipedia.org/zh/\$term	1 汶川大地震死难学生调查活动

continued on next page...

...continued from previous page

Term	GFW Rule	Pages Affected
查活动		
<i>investigation of student casualty in the Wenchuan earthquake, initiated by artist Ai Weiwei in Dec 2008.</i>		
08宪章	broad.standard	1
08憲章	broad.standard	0
〇八宪章	prefix.standard	1
零八宪章	wikipedia & \$term	3-EN:1 Talk:零八宪章 Category:零八宪章签署者 零八宪章(en)
零八憲章	wikipedia & \$term	2-EN:1 Category:零八憲章 零八憲章(en)
Charter_08	en.wikipedia.org/wiki/\$term	0-EN:1 Charter_08(en)
<i>various terms for Charter 08, a manifesto initiated by the Nobel Peace Prize laureate Liu Xiaobo.</i>		
杨佳袭警案	prefix.standard	1
<i>Yang Jia Incident. Yang Jia assaulted a Shanghai police department on Jul 1, 2008, causing six deaths.</i>		
蚁力神	broad.standard	5 蚁力神事件 蚁力神天玺集团 Talk:蚁力神 Talk:蚁力神事件
蟻力神	broad.standard	2 2007年瀋陽蟻力神群體性事件
<i>Yilishen Incident, a large scale scam in Shenyang, Liaoning Province, resulting in large scale protest in 2007.</i>		
临沂计划生育事件	prefix.standard	1
<i>An incident in Linyi, Shandong Province caused by brutality in enforcing one-child policy. Chen Guangcheng organized the opposition and was sentenced to jail.</i>		
Events before 2000		
北京之春	broad.standard	1
<i>Beijing Spring, a short period of relative political openness and freedom in China, from 1977 to 1978.</i>		
西单墙	prefix.standard	1
西单民主墙	broad.standard	2 Talk:西单民主墙
西單民主牆	broad.standard	1
民主牆運動	prefix.standard	1
<i>various terms for the “Xidan Democracy Wall”, a location in Beijing where people expressed their political views during the Beijing Spring.</i>		
布拉格之春	prefix.standard	1
<i>Prague Spring, a 1968 reform movement in Czechoslovakia.</i>		
三年自然灾害	prefix.standard	1
<i>“Three Years of Natural Disasters”, this is the official term for the Great Famine (1958-1961) caused by the Great Leap Forward.</i>		
伊宁事变	broad.standard	2 Talk:伊宁事变
伊寧事變	broad.standard	0
<i>Yining Incident, a.k.a. Three Districts Revolution, Three Districts Revolt, a Soviet-backed revolt in 1944 seeking independence of Xinjiang. Chinese authority’s positive portraying of this incident is at odds with its current iron fist approach to Uyghur affairs.</i>		
元祐更化	zh.wikipedia.org/zh-cn/\$term	1 元祐更化
<i>A historical event circa 1085, during the Song Dynasty, in which the Wang Anshi Reform was abolished. This page’s content has nothing related to China’s current affairs.</i>		

Table 8.1: GFW Rulebook: Events

8.2 Media, Publications, Censorship and Circumvention

This section covers censorship of media and press. The terms are grouped into:

- Media, newspapers and magazines.
- Publications, in reverse chronological order (by date of publication). There are 9 books and 2 TV series (not counting those related to Tibet, Uyghur and Tiananmen).
- Censorship and censorship circumvention. It is noteworthy that GFW's Wikipedia rulebook does not cover the term “防火长城” or its English term (“Great Firewall”), or alternative terms, e.g. “中国国家防火墙” (China's National Firewall). Due to GFW's HTTP response filtering (explained in Section 5, Chinese users can not really access these pages successfully, because they contain strings that are on GFW's HTTP response filtering rulebook, e.g. 法轮功 (Falungong) and/or 达赖喇嘛 (Dalai Lama).

Also worth noting is that even though the term “Great Firewall” is not on GFW rulebook, the benign term “Great Wall of China” is. More precisely, the string “en.wikipedia.org/wiki/Great_Wall_of_China” would trigger GFW reset. Isn't this puzzling?

Term	GFW Rule	Pages Affected
Media, Newspapers & Magazines		
推特	prefix.standard	2 推特中文圈
Twitter	zh.wikipedia.org/zh/\$term zh.wikipedia.org/zh-hk/\$term	3 Twitter Twitterrific Twitter服务与应用程序列表
<i>Twitter website is blocked in China, here we list the three GFW rules regarding Wikipedia for Twitter.</i>		
博讯	prefix.standard zh.wikipedia.org/zh-tw/\$term zh.wikipedia.org/zh-hant/\$term	3 博讯新闻网 博讯网
<i>Boxun, a news resource for China related issues.</i>		
多维	broad.standard	84 卢多维西 吉米多维奇 多维尔 多维尺度 多维尺度分析 多维新闻网 User_talk:多维思考 Category:多倫多维基人 ...
多維	broad.standard	22 多維新聞 埃爾多維奧 ...
<i>Dwnews, a news resource for China related issues. The two characters happen to be in Chinese phonetic translation of many foreign terms, e.g. Devillers (多维尔), hence affecting a large number of Wikipedia pages.</i>		
牛博网	prefix.standard	1
<i>Bullogger.com, a blogging website created by Luo Yonghao, which hosts many censored blog posts.</i>		
美国之音	broad.standard	3 Talk:美国之音 Category:含有来自美国之音内容的维基百科条目
美國之音	broad.standard	1
<i>Voice of America.</i>		
彭博公司	prefix.standard	1
彭博新聞社	zh.wikipedia.org/zh/\$term	1 彭博新聞社
<i>Bloomberg company, Bloomberg News Agency, blocked due to its reporting on wealth of Xi Jinping's relatives.</i>		
中国数字时代	zh.wikipedia.org/zh-cn/\$term	1 中国数字时代
<i>China Digital Times, a news website operated by Xiao Qiang et al, who is affiliated with UC Berkeley.</i>		
新世纪出版社	prefix.standard	1
<i>New Century Press, a Hong Kong based publisher, which published many books banned by China authority.</i>		
纽约时报	zh.wikipedia.org/zh-cn/\$term	10 纽约时报 纽约时报公司诉萨利文案 纽约时报畅销书榜 纽约时报广场 纽约时报报道温家宝家族财富事件 ...
<i>New York Times, blocked due to its reporting on Wen Jiabao's family wealth.</i>		
人民日报	prefix.standard zh.wikipedia.org/zh-cn/\$term	4 人民日报号外 人民日报社 人民日报藏文版
<i>People Daily, the official mouthpiece of Chinese government.</i>		
亚洲周刊	prefix.standard	1

continued on next page...

...continued from previous page		
Term	GFW Rule	Pages Affected
亞洲週刊 <i>Asia Weekly, a Hong Kong-based newsweekly.</i>	zh.wikipedia.org/zh-cn/\$term zh.wikipedia.org/zh-cn/\$term	2 亞洲週刊 亞洲週刊中文十大好書
阳光卫视 <i>Sun TV (literally sunshine satellite TV), a Chinese satellite TV station based in Hong Kong.</i>	prefix.standard	1
阳光时务 陽光時務	zh.wikipedia.org/zh-cn/\$term prefix.standard zh.wikipedia.org/zh/\$term zh.wikipedia.org/zh-cn/\$term	1 阳光时务周刊 2 陽光時務週刊
阳光时务周刊 <i>Sun Affairs, a weekly magazine based in Hong Kong.</i>	prefix.standard	1
南方都市报 <i>Southern Metropolis Daily, usually regarded as a outspoken newspaper in mainland China. Its sister newspaper Southern Weekly, had a major conflict with the propaganda authority in Jan 2013, i.e. the Southern Weekly Incident.</i>	prefix.standard zh.wikipedia.org/zh-tw/\$term zh.wikipedia.org/zh-hk/\$term zh.wikipedia.org/zh-sg/\$term zh.wikipedia.org/zh-hans/\$term	3 南方都市报事件 南方都市报案
开放杂志 開放雜誌 <i>Open Magazine, a Hong Kong based monthly magazine.</i>	broad.standard broad.standard	1 1
世界经济导报 世界經濟導報 <i>World Economic Herald, a pro-reform newspaper in the 1980's, it was shut down after the Tiananmen Incident.</i>	broad.standard broad.standard	2 Talk:世界经济导报 1
华夏文摘 華夏文摘 <i>China News Digest, a news website operated by Chinese diaspora in US.</i>	broad.standard broad.standard	1 1
自由中国论坛 自由中國論壇 <i>Free China Forum, a defunct web forum. A Wikipedia article with this title was deleted and no longer exists.</i>	broad.standard broad.standard	0 0
大参考 (网络杂志) <i>Macro Reference, a web magazine no longer in operation. It was run by 李洪宽 (Li Hongkuan).</i>	broad.standard	1
Publications		
河殤 河殤 <i>River Elegy, a documentary TV series on China history and culture, aired in 1988, banned after 1989.</i>	broad.standard broad.standard	2 Talk:河殤 0
Hu Jintao Biography <i>Chinese title is 胡锦涛传, a book by Wen Siyong and Ren Zhichu, published in 2003.</i>	zh.wikipedia.org/zh-cn/\$term zh.wikipedia.org/zh-tw/\$term zh.wikipedia.org/zh-hans/\$term	1 Hu Jintao Biography
晚年周恩来 <i>Zhou Enlai's Later Years, a banned book by Gao Wenqian published in 2003. Zhou Enlai was the first premier of People's Republic of China.</i>	zh.wikipedia.org/zh-cn/\$term	1 晚年周恩来
毛泽东私人医生回忆录 <i>Memoir of Mao Zedong's Private Doctor, a memoir by Li Zhisui, published in 2005. English title is The Private Life of Chairman Mao.</i>	zh.wikipedia.org/zh-cn/\$term	1 毛泽东私人医生回忆录
毛澤東：鮮為人知的故事 <i>Mao Zedong: the Unknown Story, a banned book by Zhang Rong and Jon Halliday, published in 2005.</i>	prefix.standard	1
激流中国 <i>Torrents in China, an award-winning documentary series by Japan Broadcasting Corporation, aired from 2007 to 2008.</i>	prefix.standard	1
维权诗集 <i>Collection of Rights Defending Poems, a banned poetry book published in 2008.</i>	prefix.standard	1
改革历程 改革歷程	broad.standard prefix.standard	2 Talk:改革历程 24

continued on next page...

...continued from previous page		
Term	GFW Rule	Pages Affected
	zh.wikipedia.org/zh-cn/\$term	
<i>Journey of Reform, a banned book published in 2009. English title is Prisoner of the State, the Secret Journal of Premier Zhao Ziyang.</i>		
大江大海一九四九	zh.wikipedia.org/zh/\$term	1 大江大海一九四九
<i>Rivers and Seas in 1949, a bestselling history book by Taiwanese author Long Yingtai, published in 2009.</i>		
谁是新中国	prefix.standard	1
<i>Who is the New China, a banned book by historian Xin Haonian, the Chinese version was published in 2012.</i>		
陈一谘回忆录	zh.wikipedia.org/zh/\$term	1 陈一谘回忆录
<i>Memoir of Chen Yizi, a banned book published in 2013.</i>		
Censorship		
金盾	zh.wikipedia.org/zh-hans/\$term	4 金盾 金盾圓龜金花蟲 金盾工程 金盾暗沙
金盾工程	broad.standard	2 Talk:金盾工程
<i>Golden Shield, a surveillance project operated by the Ministry of Public Security, which is different to, but often confused with GFW by observers.</i>		
绿坝娘	prefix.standard	1
<i>Green Dam Lady, an anthropomorphism name in response to the release of Green Dam, a government-developed content control software.</i>		
中国网络软件过滤关键字列表	broad.standard	1
<i>list of filtered keywords by Chinese internet software.</i>		
中华人民共和国被封锁网站列表	prefix.standard	1
<i>list of websites blocked by the People's Republic of China.</i>		
中华人民共和国被禁出版物列表	prefix.standard	1
<i>list of publications banned by the People's Republic of China.</i>		
Censorship Circumvention		
TOR	zh.wikipedia.org/zh-cn/\$term	43 TOR TORCS ToRo Torchlight Torixoréu Tornado Toronto Torrent Torrent文件 Tortoise_坦克 Tor函子 ...
<i>Tor, an online anonymity software.</i>		
无界	prefix.standard	4 无界浏览 无界漫游 无界网络
<i>Ultrasurf, a popular GFW-circumvention proxy software.</i>		
自由門	broad.standard	0
<i>Freemate, a popular GFW-circumvention proxy software. The simplified Chinese version is not on GFW rulebook for Wikipedia.</i>		
世界通	prefix.standard	5 世界通史 世界通用疾病分類手冊 世界通訊 世界通訊公司
<i>GPass, a deprecated GFW-circumvention software.</i>		
花园网	broad.standard	1
花園網	broad.standard	0
<i>Garden Network, a deprecated GFW-circumvention software.</i>		
西厢计划	prefix.standard zh.wikipedia.org/zh/\$term zh.wikipedia.org/zh-cn/\$term	1
<i>Xi Xiang Project, a GFW-circumvention software/project.</i>		
破网	prefix.standard	3 破网技术 破网软件
突破封锁软件	zh.wikipedia.org/zh-cn/\$term	1 突破封锁软件
突破网络审查	prefix.standard	2 突破网络审查技术
突破網絡審查	zh.wikipedia.org/zh-hant/\$term	1
<i>various terms for GFW-circumvention technology/software, internet censorship circumvention.</i>		

Table 8.2: GFW Rulebook: Media, Publications, Censorship and Circumvention

8.3 Organizations, Political & Government

This section covers:

- Organizations (NGO's and dissident organizations).
- Political terms.
- Government-related terms.

Term	GFW Rule	Pages Affected
Organizations		
无国界记者	broad.standard	2 Talk:无国界记者
無國界記者	broad.standard	3 無國界記者組織 Talk:無國界記者
<i>Reporters Without Borders.</i>		
泛蓝联盟	broad.standard	3 中国泛蓝联盟 Talk:中国泛蓝联盟 User:talk:中国泛蓝联盟
泛藍聯盟	broad.standard	1
<i>China Pan-Blue Union, a political organization in mainland China established in 2004. Several organizers were sentenced to jail.</i>		
中国民主党	broad.standard	9 中国民主党_(1988) 中国民主党派 Talk:中国民主党 Template:中国民主党派 Category:中国民主党派 ...
中國民主黨	broad.standard	9 中國民主黨_(中國大陸) 中國民主黨_(中華民國) 中國民主黨_(消歧義) 中國民主黨_(臺灣) ...
<i>China Democratic Party, a political organization established in 1998. Many organizers were sentenced to jail.</i>		
独立中文笔会	broad.standard	2 Talk:独立中文笔会
獨立中文筆會	broad.standard	0
<i>Independent Chinese PEN Center, an organization of independent Chinese authors established in 2001.</i>		
中国家庭教会	prefix.standard	1
	zh.wikipedia.org/zh-hk/\$term	
<i>China Family Church, a.k.a. China Underground Church, churches independent from China authority.</i>		
中国劳工观察	broad.standard	2 李强_(中国劳工观察组织)
中國勞工觀察	broad.standard	1
<i>China Labor Watch.</i>		
中文网志年会	prefix.standard	1
	zh.wikipedia.org/zh-hk/\$term	
	zh.wikipedia.org/zh-hant/\$term	
<i>Chinese Blogger Conference, an annual “unconference” for Chinese bloggers, started in 2005. It was cancelled in 2010 due to interference from the authority. In 2012, it was moved to “the cloud”.</i>		
中国天网人权事务中心	prefix.standard	1
<i>China Tianwang Human Rights Center, a rights defending organization in China, established in 2006.</i>		
六四天网	broad.standard	1
六四天網	broad.standard	0
<i>June 4th Tianwang, a website operated by China Tianwang Human Rights Center.</i>		
公盟法律研究中心	prefix.standard	1
<i>Open Constitution Initiative, Law Research Center. The largest NGO in China on civil rights activism. Established by Xu Zhiyong et al in 2003. After being shut down by the authority in 2009, the staff changed the name to “Citizens” and continue the work.</i>		
Political		
新闻自由	zh.wikipedia.org/zh-cn/\$term	5 新闻自由 新闻自由奖 新闻自由委员会 新闻自由指数 新闻自由掠夺者
<i>freedom of press.</i>		
和平演变	prefix.standard	1
<i>peaceful transition, a term used by China authority to describe efforts by foreign powers to subvert the Communist Party rule.</i>		
公共知识分子	prefix.standard	1
<i>public intellectual.</i>		

continued on next page...

...continued from previous page		
Term	GFW Rule	Pages Affected
敏感人士 <i>sensitive persons.</i>	broad.standard	1
秦城监狱 <i>Qincheng Prison, well known for political prisoners and high rank government officials.</i>	prefix.standard	2 秦城监狱囚犯列表
文字狱	prefix.standard	1
文字獄	prefix.standard zh.wikipedia.org/zh-cn/\$term zh.wikipedia.org/zh-hant/\$term	1
<i>jailing people for their writings and speeches.</i>		
绝食	broad.standard	7 1981年爱尔兰绝食抗议 2006年中国维权绝食接力 2006年绝食接力 2006年维权绝食接力事件 爱尔兰绝食抗议 ...
絕食 <i>hunger strike; fasting.</i>	broad.standard	15 2006年中國維權絕食接力 岩灣絕食事件 Talk:絕食 Category:死於絕食的中國人 ...
中国人权	broad.standard	13 中国人权_(组织) 中国人权状况 中国人权运动 中国人权问题 Talk:中国人权运动 Template:中国人权律师 ...
中國人權 <i>China human rights.</i>	broad.standard	9 中國人權_(組織) 中國人權協會 中國人權狀況 中國人權運動 台灣關懷中國人權聯盟 ...
中国民主运动	broad.standard	1
中國民主運動 <i>China democracy movement.</i>	broad.standard	3 Category:中國民主運動 Category:中國民主運動人物
中国持不同政见者列表	prefix.standard	1
中國持不同政見者列表	zh.wikipedia.org/zh-tw/\$term zh.wikipedia.org/zh-sg/\$term zh.wikipedia.org/zh-hant/\$term	1 中國持不同政見者列表
中國持不同政見者名單	prefix.standard zh.wikipedia.org/zh-cn/\$term zh.wikipedia.org/zh-tw/\$term zh.wikipedia.org/zh-hant/\$term	1
中華人民共和國持不同政見者列	zh.wikipedia.org/zh-tw/\$term	1 中華人民共和國持不同政見者列表
中國政治異見人士 <i>list of Chinese dissidents.</i>	prefix.standard	1
Government		
中國共產黨 <i>Chinese Communist Party (CCP), in traditional Chinese.</i>	broad.standard	92 中國共產黨領導人 中國共產黨章程 Talk:中國共產黨領導 Talk:集體 User:中國共產黨網路特務入侵中文維基百科 ...
中共五大书记 <i>This term means "five secretaries of CCP", referring to Mao Zedong, Zhou Enlai, Liu Shaoqi, Ren Bishi and Zhu De, elected in 1945.</i>	prefix.standard	1
共匪 <i>communist bandits, a derogative term for CCP.</i>	prefix.standard	1
團派 <i>Faction of the Youth League, a major faction in CCP power struggle, composed of Party cadres with leadership experience in the Youth League.</i>	zh.wikipedia.org/zh/\$term	1 團派
团中央	prefix.standard	1
共青团中央 <i>Central Committee of the Communist Youth League.</i>	prefix.standard	1
中国共产主义青年团 <i>Chinese Communist Youth League.</i>	prefix.standard	4 中国共产主义青年团中央委员会 中国共产主义青年团中央机关旧址 中国共产主义青年团团歌
中國共產主義青年團歷任領導人	prefix.standard	1

...continued from previous page		
Term	GFW Rule	Pages Affected
<i>past leaders of the Chinese Communist Youth League.</i>		
中国新民主主义青年团	zh.wikipedia.org/zh-hant/\$term	1 中国新民主主义青年团
<i>Chinese New Democracy Youth League, a precursor to the Chinese Communist Youth League.</i>		
胡温体制	prefix.standard	1
<i>Hu-Wen System, Hu and Wen refer to Hu Jintao and Wen Jiabao, president and premier of China from 2002 to 2012.</i>		
社会主义初级阶段	zh.wikipedia.org/zh-hant/\$term	2 社会主义初级阶段 社会主义初级阶段理论
<i>"Primary Stage of Socialism", a terminology in CCP theories.</i>		
一个中心	prefix.standard	3 一个中心、两个基本点 一个中心两个基本点
<i>"One Central Task", combined with "两个基本点(two basic points)", is CCP's grand strategy, the central task is economic development, the two basic points are 1. stick to four core principles, 2. stick to reform and opening-up.</i>		
太子党	prefix.standard	1
<i>Princeling</i>		
Crown Prince Party	zh.wikipedia.org/zh/\$term zh.wikipedia.org/zh-cn/\$term .wikipedia.org/wiki/\$term en.wikipedia.org/wiki/\$term	0-EN:2 Princeling(en) Princelings(en) 0-EN:1 Crown.Prince.Party(en)
<i>descendants of top Party officials in China, many of whom holds high-level political and business positions.</i>		
网络评论员	prefix.standard zh.wikipedia.org/zh-cn/\$term zh.wikipedia.org/zh-tw/\$term	2 网络评论员_(广义)
<i>web commentators organized and/or sponsored by the government.</i>		
中共中央宣傳部	broad.standard	1
<i>Propaganda Department of the Central Committee of CCP.</i>		
妇联	zh.wikipedia.org/zh-cn/\$term	1 妇联
中华全国妇女联合会	prefix.standard	1
<i>All-China Women's Federation. The term is not sensitive, but its Wikipedia page contains a June 4th Incident template, which was added on Oct 23, 2011. This suggests that GFW added this rule after this date.</i>		
中华全国归国华侨联合会	prefix.standard	1
<i>All-China Federation of Returned Overseas. Non-sensitive term, same story as the previous entry.</i>		
中国法学会	zh.wikipedia.org/zh-hk/\$term	2 中国法学会 中国法学会宪法学研究会
<i>China Law Society. Non-sensitive term, same story as the previous two entries.</i>		
中央警卫局	prefix.standard	1
中央警衛局	prefix.standard	1
<i>CCP Central Guard Bureau.</i>		
国家国防动员委员会	prefix.standard	1
<i>National Defense Mobilization Commission.</i>		
中国人民解放军总后勤部	prefix.standard	3 中国人民解放军总后勤部军事交通运输部 中国人民解放军
<i>Chinese People's Liberation Army General Logistics Department, one of four General Departments of PLA. The other three are not blocked: General Staff Department (总参谋部), General Political Department (总政治部), General Armaments Department (总装备部).</i>		
中华人民共和国贪污案件	zh.wikipedia.org/zh-cn/\$term	2 中华人民共和国贪污案件 中华人民共和国贪污案件列表
中華人民共和國貪污案件	prefix.standard	1
<i>list of corruption cases in People's Republic of China.</i>		
2012年中华人民共和国腐败案件	prefix.standard	1
<i>list of corruption cases in People's Republic of China in 2012. Other years are not blocked.</i>		
中华人民共和国宪法	zh.wikipedia.org/zh-tw/\$term prefix.standard	2 中华人民共和国宪法历史
<i>Constitution of People's Republic of China.</i>		

Table 8.3: GFW Rulebook: Organizations, Political & Government

8.4 People

This section covers people names, but excluding those related to Hong Kong, Taiwan, Tibet, Uyghur, Falungong and five terms in the SELF category, (see respective sections). I group these names into the following subcategories:

- Dissidents, further divided into:
 - “Newer” terms (see explanation below).
 - “Older” terms (see explanation below).
- Government officials, further divided into:
 - Names (and nicknames) of the 9 members of the 17th Politburo Standing Committee, and their relatives.
 - Names of the 15 members of the 17th Politburo not in the Standing Committee.
 - Names of past top Party leaders.
 - Other lower-level government officials, whose names are blocked either due to scandals or their close relationship to the censorship apparatus.
- Misc, a few names who are neither dissidents, nor government officials.

For dissidents, I observe two types. One type seems to be “newer”, mainly by prefix rules. The other type seems to be “older”, and are blocked by broad rules, furthermore, both the simplified and traditional versions are blocked (by broad rules), even for cases when Wikipedia does not have an article for one version. My hypothesis for the “older” version is that the censorship official handed a (long) list of names to the GFW operator (probably just in one “transaction”), who did the due diligence and blocked both the simplified and traditional versions using the broad rule. For the “newer” type, it is likely the censorship officials were giving URLs to the GFW operators, presumably in many transactions.

Note that a few people likely got both - i.e., they are in that “broad rule” list, but GFW “revisited” them on alternative names (or Wikipedia titles). These cases are: 丁子霖, 侯德建, 王丹, 唯色, 赵紫阳. I list them in the “Newer” section.

The rationale to divide the Officials category into these four groups is that GFW rulebook has standard treatment to names of the members of the 17th Politburo (in office between 2007 to 2012). For all of these 25 names, they are blocked by the broad rule, and have both the simplified and traditional versions blocked. For the newer batch (i.e. the 18th Politburo from 2012) or older batches (16th Politburo or earlier), there is no such standard treatment.

Term	GFW Rule	Pages Affected
Dissidents (<i>newer</i> terms)		
艾未未	prefix.standard zh.wikipedia.org/zh/\$term zh.wikipedia.org/zh-cn/\$term zh.wikipedia.org/zh-tw/\$term	2 艾未未星
<i>Ai Weiwei, dissident artist.</i>		
包遵信	zh.wikipedia.org/zh/\$term zh.wikipedia.org/zh-cn/\$term zh.wikipedia.org/zh-tw/\$term zh.wikipedia.org/zh-hk/\$term zh.wikipedia.org/zh-sg/\$term	1 包遵信
<i>Bao Zunxin, (1937 - 2007/10/28), prominent dissident scholar.</i>		
北岛_(诗人)	zh.wikipedia.org/zh-tw/\$term	1 北岛_(诗人)
<i>Bei Dao, dissident poet, relocated back to Hong Kong in 2007.</i>		
陈光诚	prefix.standard zh.wikipedia.org/zh/\$term zh.wikipedia.org/zh-hans/\$term	1
陳光誠	prefix.standard	2 陳光誠訪美事件

continued on next page...

...continued from previous page		
Term	GFW Rule	Pages Affected
Chen.Guangcheng	en.wikipedia.org/wiki/\$term	0-EN:1 Chen.Guangcheng(en)
<i>Chen Guangcheng, the legendary blind civil rights lawyer/activist, who escaped from house arrest and fled to U.S. Embassy in Beijing in April 2012.</i>		
陳一諮	zh.wikipedia.org/zh/\$term	2 陳一諮 陳一諮回憶錄
<i>Chen Yizi, dissident economist, who published a memoir in May 2013. The simplified version of his name is not on GFW rulebook.</i>		
程翔	prefix.standard	1
<i>Cheng Xiang, journalist, arrested for espionage in Apr 2005, released in Feb 2008.</i>		
丁子霖	broad.standard	2 Talk:丁子霖
	zh-yue.wikipedia.org/wiki/\$term	
<i>Ding Zilin, organizer of Tiananmen Mothers, a group of relatives of victims in the June 4th Incident. One of only four terms blocked on the Yue (Cantonese) Wikipedia.</i>		
杜导正	zh.wikipedia.org/zh/\$term	1 杜导正
	zh.wikipedia.org/zh-tw/\$term	
	zh.wikipedia.org/zh-sg/\$term	
<i>Du Daozheng, pro-reform Party veteran, president of the magazine Yan Huang Chun Qiu.</i>		
冯正虎	zh.wikipedia.org/zh-cn/\$term	1 冯正虎
	zh.wikipedia.org/zh-hk/\$term	
<i>Feng Zhenghu, activist. Feng was refused re-entry into China in 2009, he protested and remained in Narita International Airport (Tokyo) for 92 days, which attracted worldwide media attention.</i>		
侯德健	prefix.standard	1
侯德建	broad.standard	1
<i>Hou Dejian, Taiwanese musician, June 4th Movement participant. The first is his correct name, the second is a misspelling.</i>		
賈甲	zh.wikipedia.org/zh-cn/\$term	1 賈甲
<i>Jia Jia, dissident, father of Jia Kuo (see below).</i>		
贾阔	prefix.standard	1
<i>Jia Kuo, dissident, son of Jia Jia (see above).</i>		
姜维平	prefix.standard	1
	zh.wikipedia.org/zh/\$term	
<i>Jiang Weiping, journalist, jailed for six years for reporting secrets of Bo Xilai.</i>		
蒋彦永	zh.wikipedia.org/zh-cn/\$term	1 蒋彦永
	zh.wikipedia.org/zh-tw/\$term	
	zh.wikipedia.org/zh-sg/\$term	
	zh.wikipedia.org/zh-hans/\$term	
<i>Jiang Yanyong, a high ranking doctor in PLA medical system. He leaked the truth about SARS pandemic to the media in Apr 2003.</i>		
李旺陽	prefix.standard	3 李旺陽事件 李旺陽被自殺事件
	zh.wikipedia.org/zh-cn/\$term	
<i>Li Wangyang, (1950 - 2012/06/06), an activist who spent 22 years in jail for organizing protest in 1989. His sudden death in Jun 2012 was claimed to be suicide by China authority, which convinced few people. This incident raised major attention in Hong Kong.</i>		
梁海怡	prefix.standard	1
<i>Liang Haiyi, activist, jailed for her activism in Chinese Jasmine Revolution.</i>		
廖亦武	prefix.standard	1
<i>Liao Yiwu, dissident poet and writer. He fled China in Jul 2011.</i>		
劉霞_(劉曉波妻)	zh.wikipedia.org/zh-sg/\$term	1 劉霞_(劉曉波妻)
	zh.wikipedia.org/zh-hans/\$term	
<i>Liu Xia, artist, wife of Liu Xiaobo.</i>		
刘晓波	broad.standard	2 Talk:刘晓波
	en.wikipedia.org/wiki/\$term	
劉曉波	en.wikipedia.org/wiki/\$term	0-EN:1 劉曉波(en)
	zh.wikipedia.org/zh-hk/\$term	
	zh.wikipedia.org/zh-tw/\$term	
	zh-yue.wikipedia.org/wiki/\$term	

...continued from previous page		
Term	GFW Rule	Pages Affected
Liu_Xiaobo	en.wikipedia.org/wiki/\$term	0-EN:4 Liu_Xiaobo(en) Liu_Xiaobo_(intellectual)(en) Liu_Xiaobo_(taekwondo)(en) Liu_xiaobo(en)
<i>Liu Xiaobo, prominent dissident scholar; 2010 Nobel Peace Prize laureate.</i>		
Liu_Xianbin	en.wikipedia.org/wiki/\$term	0-EN:2 Liu_Xianbin(en) Liu_xianbin(en)
<i>dissident. He was sentenced to jail three times, the last one in Mar 2011. Surprisingly, the Chinese terms 刘贤斌/劉賢斌 are not on GFW rulebook.</i>		
刘逸明	prefix.standard	1
<i>Liu Yiming, author; arrested for allegedly spreading rumors in the “70 Kmph Incident” in May 2009.</i>		
盤古樂團	prefix.standard zh.wikipedia.org/zh-tw/\$term gan.wikipedia.org/wiki/\$term	1
<i>Pangu Band, an avant-garde punk band. This term is the only one targeting the Gan (Jiangxi dialect) Wikipedia. A related term “盘古乐队” (in simplified Chinese) is blocked on its own.</i>		
浦志强	prefix.standard zh.wikipedia.org/zh-sg/\$term zh.wikipedia.org/zh-hans/\$term	1
<i>Pu Zhiqiang, rights defending lawyer.</i>		
冉云飞	prefix.standard	1
<i>Ran Yunfei, writer, detained briefly in Chinese Jasmine Revolution.</i>		
孙文广	prefix.standard	1
<i>Sun Wenguang, dissident scholar.</i>		
谭作人	zh.wikipedia.org/zh-cn/\$term	1 谭作人
<i>Tan Zuoren, activist, sentenced to jail in Feb 2010, for investigating tofu-dregs schoolhouses in the Sichuan Earthquake.</i>		
唐柏桥	zh.wikipedia.org/zh-tw/\$term	1 唐柏桥
<i>Tang Baiqiao, dissident.</i>		
王丹	prefix.standard	9 王丹_(1969年) 王丹_(東漢) 王丹_(民运人士) 王丹_(民運人士) 王丹_(消歧義) 王丹_(游泳运动员) 王丹凤 王丹桂
	zh.wikipedia.org/zh/\$term zh.wikipedia.org/zh-sg/\$term	
王丹_(民运人士)	broad.standard	1
王丹_(民運人士)	broad.standard	1
<i>Wang Dan, student leader in the Tiananmen Protest in 1989.</i>		
唯色	broad.standard	3 茨仁唯色 Talk:唯色
程文萨	prefix.standard	1
<i>Tsering Woesser, Tibetan writer. 程文萨 is an obscure alternate name of 唯色.</i>		
夏业良	prefix.standard	1
<i>Xia Yeliang, dissident, who published an open letter to Liu Yunshan, then-Party Propaganda Chief in May 2009.</i>		
许志永	zh.wikipedia.org/zh-cn/\$term	1 许志永
<i>Xu Zhiyong, prominent legal scholar and activist, an iconic figure in China's New Citizen Movement.</i>		
阎明复	broad.standard	2 Talk:阎明复
<i>Yan Mingfu, Party veteran, he was removed from his posts after the Tiananmen crackdown.</i>		
杨银波	prefix.standard	1
<i>Yang Yinbo, a “post-80” writer, the youngest entry in GFW's Wikipedia rulebook.</i>		
姚监复	prefix.standard zh.wikipedia.org/zh/\$term zh.wikipedia.org/zh-cn/\$term zh.wikipedia.org/zh-hans/\$term	1
<i>Yao Jianfu, scholar; authored memoir of Chen Xitong, the jailed former Beijing Party head, in May 2012.</i>		
张伯笠	prefix.standard	1
<i>Zhang Boli, a leader in the Tiananmen Protest in 1989. He fled China in 1991, later converted to Christianity and became a popular pastor.</i>		

...continued from previous page		
Term	GFW Rule	Pages Affected
張世軍	prefix.standard zh.wikipedia.org/zh-sg/\$term zh.wikipedia.org/zh-hans/\$term	1
<i>Zhang Shijun, a PLA officer experienced the Tiananmen crackdown in 1989. He was the first PLA officer to publicly denounce the crackdown. He published an open letter to Hu Jintao using real name in Mar 2009.</i>		
赵连海	zh.wikipedia.org/zh-cn/\$term	1 赵连海
<i>Zhao Lianhai, activist, organizer of "Home for Kidney Stone Babies", a group of family of victims in the melamine-tainted milk scandal in 2008.</i>		
赵紫阳	broad.standard	5 Talk:赵紫阳 User_talk:赵紫阳说我帅 Category:赵紫阳 Category:赵紫阳相关书籍
趙紫陽	broad.standard	3 趙紫陽軟禁中的談話 Talk:趙紫陽
赵修业	prefix.standard	1
<i>Zhao Ziyang, (1919 - 2005/01/17), former premier and CCP general secretary, a pro-reform politician who was removed from post in 1989 and remained under house arrest until his death in 2005. 赵修业 (Zhao Xiuye) is an obscure alternate name.</i>		
宗鳳鳴	prefix.standard	1
<i>Zong Fengming, (1920 - 2910/01/07), Party veteran, author of Interviews with Zhao Ziyang, a book published in Jan 2007 in Hong Kong.</i>		
Dissidents (older terms)		
鲍彤	broad.standard	2 Talk:鲍彤
鮑彤	broad.standard	1
<i>Bao Tong, Party veteran, secretary of Zhao Ziyang.</i>		
曹长青	broad.standard	1 User:曹长青
曹長青	broad.standard	2 Talk:曹長青
<i>Cao Changqing, dissident writer.</i>		
柴玲	broad.standard	2 Talk:柴玲
<i>Chai Ling, student leader of the Tiananmen Protest in 1989.</i>		
陈奎德	broad.standard	1
陳奎德	broad.standard	2 Talk:陳奎德
<i>Chen Kuide, dissident scholar.</i>		
陈子明	broad.standard	3 Talk:陈子明 User_talk:陈子明
陳子明	broad.standard	1
<i>Chen Ziming, dissident.</i>		
方励之	broad.standard	2 Talk:方励之
方勵之	broad.standard	2 Talk:方勵之
<i>Fang Lizhi, (1936 - 2012/04/06), dissident scholar, iconic figure of Tiananmen Protest in 1989 and Chinese democracy movement.</i>		
方圓	broad.standard	15 方圓_(1948年) 方圓庙 方圓街道 活步方圓架 User:方圓圓 User_talk:方圓 User_talk:方圓之子 User_talk:方圓僧 ...
方圓	broad.standard	5 方圓社 方圓美術館 Talk:方圓社 Talk:方圓美術館 User_talk:方圓之間
<i>Fang Yuan, dissident. His name means "square" and "circle", which appears in many phrases in Chinese.</i>		
封从德	broad.standard	2 Talk:封从德
封從德	broad.standard	1
<i>Feng Congde, student leader in Tiananmen Protest in 1989.</i>		
高瑜	broad.standard	1
<i>Gao Yu, female dissident journalist.</i>		
高智晟	broad.standard	2 Talk:高智晟
<i>Gao Zhisheng, rights defending lawyer.</i>		
韩东方	broad.standard	1
韓東方	broad.standard	1
<i>Han Dongfang, activist.</i>		
胡佳_(1973年)	broad.standard	1
<i>Hu Jia, activist.</i>		

...continued from previous page		
Term	GFW Rule	Pages Affected
胡平_(作家) <i>Hu Ping, dissident scholar.</i>	broad.standard	2 Talk:胡平_(作家)
黄琦 黃琦 <i>Huang Qi, activist.</i>	broad.standard broad.standard	2 Talk:黄琦 1
蒋品超 蔣品超 <i>Jiang Pinchao, dissident.</i>	broad.standard broad.standard	2 Talk:蒋品超 1
焦国标 焦國標 <i>Jiao Guobiao, dissident, former Peking University professor, published an article “Denouncing the Central Propaganda Department” in Mar 2004.</i>	broad.standard broad.standard	2 Talk:焦国标 0
黎安友 <i>Li Anyou, Chinese name of Andrew Nathan, American scholar specializes in Chinese issues.</i>	broad.standard	1
李洪宽 李洪寬 <i>Li Hongkuan, dissident.</i>	broad.standard broad.standard	1 0
李禄 李祿 <i>Li Lu, student leader in the Tiananmen Protest in 1989.</i>	broad.standard broad.standard	0 1
李慎之 <i>Li Shen zhi, (1923 – 2003/04/22), scholar, a prominent exponent of Chinese liberalism.</i>	broad.standard	2 Talk:李慎之
刘宾雁 劉賓雁 <i>Liu Binyan, (1925 - 2005/12/05), dissident, prominent figure in the Tiananmen Protest in 1989.</i>	broad.standard broad.standard	2 Talk:刘宾雁 0
刘刚 劉剛 <i>Liu Gang, student leader in the Tiananmen Protest in 1989. This is a very common Chinese name.</i>	broad.standard broad.standard	2 User.talk:刘刚 User.talk:刘刚阿訇 5 劉剛_(民運人士) 劉剛_(運動員) Talk:劉剛 ...
齐志勇 齊志勇 <i>Qi Zhiyong, activist, injured in the June 4th crackdown in 1989.</i>	broad.standard broad.standard	2 Talk:齐志勇 1
师涛 師濤 <i>Shi Tao, journalist, arrested in 2004 and jailed until Aug 23, 2013. His conviction involved Yahoo's cooperation to the China authority, causing huge damage to Yahoo reputation.</i>	broad.standard broad.standard	2 Talk:师涛 1
王炳章 <i>Wang Bingzhang, prominent activist in China's democracy movement, serving life in prison currently.</i>	broad.standard	2 王炳章_(光緒進士)
王军涛 王軍濤 <i>Wang Juntao, prominent dissident scholar.</i>	broad.standard broad.standard	1 0
王力雄 <i>Wang Lixiong, writer, husband of Woesser.</i>	broad.standard	2 Talk:王力雄
王若望 <i>Wang Ruowang, (1918 - 2001/12/19), dissident.</i>	broad.standard	2 Talk:王若望
王文怡 <i>Wang Wenyi, reporter of Epoch Times.</i>	broad.standard	2 Talk:王文怡
王有才 <i>Wang Youcai, dissident.</i>	broad.standard	2 Talk:王有才
魏京生 <i>Wei Jingsheng, dissident, the first imprisoned activist after the end of the Cultural Revolution.</i>	broad.standard	2 Talk:魏京生

...continued from previous page		
Term	GFW Rule	Pages Affected
文炎 <i>Wen Yan, a "post-80" activist, organized 中国泛蓝联盟(China Pan-Blue Union).</i>	broad.standard	3 吳文炎 Talk:文炎
吴弘达 吳弘達 <i>Wu Hongda, dissident.</i>	broad.standard broad.standard	1 2 Talk:吳弘達
吾尔开希 吾爾開希 <i>Wuer Kaixi, student leader of the Tiananmen Protest in 1989.</i>	broad.standard broad.standard	2 Talk:吾尔开希 1
辛灏年 辛灝年 <i>Xin Haonian, historian, author of the banned book 谁是新中国(Who is the New China).</i>	broad.standard broad.standard	2 Talk:辛灏年 1
熊焱 <i>Xiong Yan, student leader of the 1989 Tiananmen Protest.</i>	broad.standard	1
许家屯 許家屯 <i>Xu Jiatusun, Party veteran, was CCP's Party head in Hong Kong from 1983 to 1989, defected after the Tiananmen crackdown.</i>	broad.standard broad.standard	2 Talk:许家屯 2 Talk:許家屯
严家其 嚴家其 <i>Yan Jiaqi, prominent dissident scholar.</i>	broad.standard broad.standard	2 Talk:严家其 1
杨建利 楊建利 <i>Yang Jianli, dissident activist.</i>	broad.standard broad.standard	1 1
杨炼 楊煉 <i>Yang Lian, dissident poet and writer.</i>	broad.standard broad.standard	2 Talk:杨炼 1
姚勇战 姚勇戰 <i>Yao Yongzhan, student leader in student protest in 1989.</i>	broad.standard broad.standard	0 2 Talk:姚勇戰
余杰 <i>Yu Jie, dissident writer.</i>	broad.standard	4 Talk:余杰 User_talk:余杰刚 Wikipedia:繁体求/錯誤:余杰→余傑
袁红冰 袁紅冰 <i>Yuan Hongbing, dissident.</i>	broad.standard broad.standard	2 Talk:袁红冰 1
远志明 遠志明 <i>Yuan Zhiming, dissident writer, co-authored 河殇(River Elegy), converted to Christianity and became a popular pastor.</i>	broad.standard broad.standard	3 Talk:远志明 User_talk:远志明 1
章诒和 章詒和 <i>Zhang Yihe, writer.</i>	broad.standard broad.standard	2 Talk:章诒和 1
周国聪 周國聰 <i>Zhou Guocong, a student protester in Chengdu in 1989, who was beaten to death in police custody.</i>	broad.standard broad.standard	2 Talk:周国聪 0
Government Officials (Members of the 17th Politburo Standing Committee)		
胡锦涛 胡錦濤 <i>Hu Jintao, President of China and General Secretary of CCP from 2002 to 2012.</i>	broad.standard broad.standard	7 胡锦涛传 Talk:胡锦涛 User:胡锦涛 User_talk:胡锦涛 Category:胡锦涛 Category:胡锦涛家族 6 胡錦濤訪港 User:胡錦濤下台 User:胡錦濤是世界第一大猛種 User:胡錦濤是猛種 User_talk:630追擊胡錦濤
胡海峰 <i>Hu Haifeng, son of Hu Jintao, involved in the Nuctech-Namibia scandal.</i>	zh.wikipedia.org/zh-cn/\$term	2 胡海峰 胡海峰_(演员)

...continued from previous page

Term	GFW Rule	Pages Affected
胡海清	zh.wikipedia.org/zh/\$term zh.wikipedia.org/zh-cn/\$term zh.wikipedia.org/zh-hans/\$term zh.wikipedia.org/zh-hant/\$term	1 胡海清
<i>Hu Haiqing, daughter of Hu Jintao.</i>		
吴邦国	broad.standard	3 Talk:吴邦国 User:吴邦国
吳邦國	broad.standard	1
<i>Wu Bangguo, famous for “五不搞” (Five Don'ts) in 2011. 五不搞 is not blocked.</i>		
温家宝	broad.standard	14 中国影帝温家宝 温家宝家族财富被曝光事件 温家宝被扔鞋事件 Talk:温家宝 Talk:温家宝被扔鞋事件 ...
温家寶	broad.standard	2 中國影帝温家寶
溫總	prefix.standard	2 溫總理
温影帝	zh.wikipedia.org/zh-cn/\$term	1
	zh.wikipedia.org/zh-tw/\$term	
	zh.wikipedia.org/zh-sg/\$term	
	zh.wikipedia.org/zh-hans/\$term	
	prefix.standard	
	zh.wikipedia.org/zh-cn/\$term	
	zh.wikipedia.org/zh-tw/\$term	
zh.wikipedia.org/zh-sg/\$term		
zh.wikipedia.org/zh-hans/\$term		
<i>Wen Jiabao, Premier of China from 2002 to 2012. 溫總 is short for Premier Wen; 温影帝 is “best actor Wen”.</i>		
张培莉	prefix.standard	1
	zh.wikipedia.org/zh/\$term	
	zh.wikipedia.org/zh-cn/\$term	
	zh.wikipedia.org/zh-tw/\$term	
	zh.wikipedia.org/zh-hk/\$term	
	zh.wikipedia.org/zh-sg/\$term	
	zh.wikipedia.org/zh-hans/\$term	
	zh.wikipedia.org/zh-hant/\$term	
<i>Zhang Peili, wife of Wen Jiabao.</i>		
温云松	broad.standard	1
<i>Wen Yunsong, son of Wen Jiabao.</i>		
贾庆林	broad.standard	3 Talk:贾庆林 User:贾庆林
賈慶林	broad.standard	1
<i>Jia Qinglin.</i>		
李长春	broad.standard	2 Talk:李长春
李長春	broad.standard	1
<i>Li Changchun, Propaganda Chief from 2002 to 2012.</i>		
习近平	broad.standard	4 习近平党八条 Talk:习近平 User:习近平
習近平	broad.standard	10 習近平南巡 Talk:習近平南巡 Category:習近平 ...
<i>Xi Jinping, President of China and General Secretary of CCP from 2012.</i>		
彭丽媛	prefix.standard	1
<i>Peng Liyuan, singer, wife of Xi Jinping.</i>		
习仲勋	prefix.standard	1
<i>Xi Zhongxun, communist revolutionary, father of Xi Jinping.</i>		
习明泽	prefix.standard	1
	zh.wikipedia.org/zh/\$term	
	zh.wikipedia.org/zh-cn/\$term	
	zh.wikipedia.org/zh-tw/\$term	
	zh.wikipedia.org/zh-sg/\$term	
	zh.wikipedia.org/zh-hans/\$term	

continued on next page...

...continued from previous page		
Term	GFW Rule	Pages Affected
<i>Xi Mingze, daughter of Xi Jinping.</i>		
齊橋橋	zh.wikipedia.org/zh/\$term zh.wikipedia.org/zh-cn/\$term zh.wikipedia.org/zh-hk/\$term zh.wikipedia.org/zh-sg/\$term zh.wikipedia.org/zh-hans/\$term	1 齊橋橋
<i>Qi Qiaoqiao, Xi Jinping's sister, whose wealth was reported by Bloomberg in Jun 2012.</i>		
邓家贵	zh.wikipedia.org/zh/\$term zh.wikipedia.org/zh-cn/\$term zh.wikipedia.org/zh-tw/\$term zh.wikipedia.org/zh-hk/\$term zh.wikipedia.org/zh-sg/\$term zh.wikipedia.org/zh-hans/\$term zh.wikipedia.org/zh-hant/\$term	1 邓家贵
<i>Deng Jiagui, Xi Jinping's brother-in-law, whose wealth was reported by Bloomberg in Jun 2012.</i>		
李克强	broad.standard	3 Talk:李克强 Category:李克强
李克强	broad.standard	3 李克强訪問香港期間保安爭議 Talk:李克强訪問香港期間保安爭議
<i>Li Keqiang, Premier of China from 2012.</i>		
贺国强	broad.standard	2 Talk:贺国强
賀國強	broad.standard	1
<i>He Guoqiang.</i>		
周永康	broad.standard	6 周永康_(美國) 周永康_(美國政治人物) Talk:周永康 User:周永康 User_talk:周永康
<i>Zhou Yongkang, China's Police Chief between 2007 and 2012.</i>		
Government Officials (Members of the 17th Politburo)		
王刚	broad.standard	10 王刚_(1964年出生) 王刚_(人艺演员) 王刚_(政治人物) 王刚_(民进) 王刚_(演员) 王刚_(足球运动员) Talk:王刚_(演员)
王剛	broad.standard	...
<i>Wang Gang, this is a very common Chinese name.</i>		
王乐泉	broad.standard	2 Talk:王乐泉
王樂泉	broad.standard	1
<i>Wang Lequan.</i>		
王兆国	broad.standard	2 Talk:王兆国
王兆國	broad.standard	1
<i>Wang Zhaoguo.</i>		
王岐山	broad.standard	2 Talk:王岐山
<i>Wang Qishan.</i>		
回良玉	broad.standard	2 Talk:回良玉
<i>Hui Liangyu.</i>		
刘淇	broad.standard	2 Talk:刘淇
劉淇	broad.standard	1
<i>Liu Qi.</i>		
刘云山	broad.standard	4 Talk:刘云山 User:刘云山 User_talk:刘云山
劉雲山	broad.standard	1
<i>Liu Yunshan, director of the Propaganda Department of CCP Central Committee from 2007 to 2012, then promoted to the Standing Committee of the 18th Politburo.</i>		
刘延东	broad.standard	1
劉延東	broad.standard	1
<i>Liu Yandong.</i>		
李源潮	broad.standard	3 Talk:李源潮 User_talk:李源潮

...continued from previous page		
Term	GFW Rule	Pages Affected
<i>Li Yuanchao.</i>		
汪洋	broad.standard	25 汪洋_(将军) 汪洋_(棋手) 汪洋_(消歧义) 汪洋_(消歧義) 汪洋_(足球运动员) 汪洋中的一條船 汪洋台 汪洋泡 ...
<i>Wang Yang. The word also means “ocean”, hence very common.</i>		
张高丽	broad.standard	2 Talk:张高丽
張高麗	broad.standard	1
<i>Zhang Gaoli.</i>		
张德江	broad.standard	2 Talk:张德江
張德江	broad.standard	1
<i>Zhang Dejiang.</i>		
俞正声	broad.standard	2 Talk:俞正声
俞正聲	broad.standard	1
<i>Yu Zhengsheng.</i>		
徐才厚	broad.standard	2 Talk:徐才厚
<i>Xu Caihou.</i>		
郭伯雄	broad.standard	2 Talk:郭伯雄
<i>Guo Boxiong.</i>		
薄熙来	broad.standard	12 薄熙来之歌 薄熙来事件 薄熙来案 Talk:薄熙来 Talk:薄熙来之歌 Talk:薄熙来事件 Template:薄熙来 Category:薄熙来 ...
薄熙來	broad.standard	4 薄熙來之歌 User:薄熙來 User_talk:薄熙來
<i>Bo Xilai, ex-Party chief of Chongqing, sentenced to life in prison in Sep 2013 due to the Neil Heywood murder case and Wang Lijun Incident.</i>		
Officials (Past High Ranking Officials)		
邓小平	zh.wikipedia.org/zh-hans/\$term	10 邓小平 邓小平_(消歧义) 邓小平_(电影) 邓小平办公室 邓小平南巡 邓小平文选 邓小平时代 邓小平理论 ...
邓斌	prefix.standard zh.wikipedia.org/zh-hans/\$term	1
<i>Deng Xiaoping. 邓斌 (Deng Bin) is an obscure alternate name of Deng Xiaoping.</i>		
邓卓棣	prefix.standard zh.wikipedia.org/zh-tw/\$term	1
<i>Deng Zhuodi, grandson of Deng Xiaoping.</i>		
華國鋒	broad.standard	1
<i>Hua Guofeng, passed away on Aug 20, 2008.</i>		
万里	wikipedia.org & \$term	74-EN:2 万里_(消歧义) 万里乡 万里长城 万里长沙 万里防火墙 黄万里 Talk:万里 万里(en) 万里长城(en) ...
<i>Wan Li. The word means ten thousand (Chinese) miles, appearing in many phrases, in particular, 万里长城 (ten thousand miles of Great Wall).</i>		
江泽民	broad.standard	9 江泽民_(汽车专家) 江泽民文选 Talk:江泽民 User:江泽民 File:江泽民文选第1卷封面.jpg Category:江泽民 ...
江澤民	broad.standard	6 亞洲電視誤報江澤民逝世事件 亞視報導江澤民逝世事件 江澤民專機事件 Template:User_江澤民理論體系 ...
<i>Jiang Zemin, President of China and General Secretary of CCP from 1989 to 2002.</i>		
江绵恒	prefix.standard	1
<i>Jiang Mianheng, son of Jiang Zemin.</i>		
江绵康	zh.wikipedia.org/zh-cn/\$term	1 江绵康
<i>Jiang Miankang, son of Jiang Zemin.</i>		
李鹏	prefix.standard zh.wikipedia.org/zh/\$term zh.wikipedia.org/zh-hk/\$term zh.wikipedia.org/zh-sg/\$term	5 李鹏_(消歧义) 李鹏下台嵌字诗 李鹏图 李鹏新
李鵬	prefix.standard zh.wikipedia.org/zh-cn/\$term zh.wikipedia.org/zh-sg/\$term	4 李鵬六四日記 李鵬飛 李鵬飛_(光緒進士)

continued on next page...

...continued from previous page		
Term	GFW Rule	Pages Affected
	zh.wikipedia.org/zh-hans/\$term	
<i>Li Peng, Premier of China from 1988 to 1998.</i>		
李瑞环	prefix.standard	1
李瑞環	prefix.standard	1
<i>Li Ruihuan.</i>		
刘华清	zh.wikipedia.org/zh-tw/\$term	1 刘华清
劉華清	prefix.standard	1
	zh.wikipedia.org/zh-hk/\$term	
<i>Liu Huaqing.</i>		
曾庆红	prefix.standard	4 曾庆红_(女) 曾庆红_(消歧义) 曾庆红_(男)
	zh.wikipedia.org/zh/\$term	
	zh.wikipedia.org/zh-cn/\$term	
	zh.wikipedia.org/zh-tw/\$term	
	zh.wikipedia.org/zh-hk/\$term	
曾慶紅	prefix.standard	2 曾慶紅_(消歧義)
	zh.wikipedia.org/zh-cn/\$term	
<i>Zeng Qinghong.</i>		
Officials (other)		
曹建明	zh.wikipedia.org/zh-cn/\$term	1 曹建明
<i>Cao Jianming.</i>		
陈世炬	prefix.standard	1
	zh.wikipedia.org/zh/\$term	
	zh.wikipedia.org/zh-cn/\$term	
<i>Chen Shiju, secretary of Hu Jintao.</i>		
陳希同	prefix.standard	1
<i>Chen Xitong, ex-Party Chief of Beijing, jailed for corruption.</i>		
迟浩田	prefix.standard	1
<i>Chi Haotian.</i>		
何挺	prefix.standard	1
<i>He Ting, successor to Wang Lijun as Chongqing Police Chief.</i>		
李鸿忠	prefix.standard	1
	zh.wikipedia.org/zh-cn/\$term	
	zh.wikipedia.org/zh-hk/\$term	
	zh.wikipedia.org/zh-sg/\$term	
	zh.wikipedia.org/zh-hans/\$term	
李鴻忠	prefix.standard	1
	zh.wikipedia.org/zh-cn/\$term	
	zh.wikipedia.org/zh-tw/\$term	
<i>Li Hongzhong, Party Chief of Hubei Province, who became famous for grabbing the recording pen of a female reporter in Mar 2010.</i>		
栗智	zh.wikipedia.org/zh-tw/\$term	1 栗智
<i>Su Zhi, ex-Party Chief of Urumuqi, capital city of Xinjiang Uyghur Autonomous Region. He was removed from the post in September 2009.</i>		
令计划	prefix.standard	1
令計劃	prefix.standard	1
<i>Ling Jihua, Hu Jintao's secretary. His son died in a car crash on Mar 18, 2012, a major political scandal.</i>		
令谷	broad.standard	2 Talk:令谷
<i>Ling Gu, son of Ling Jihua, died in a car accident on Mar 18, 2012.</i>		
刘红薇	prefix.standard	1
<i>Liu Hongwei.</i>		
娄勤俭	broad.standard	1
<i>Lou Qinjian, former Deputy Minister of Industry and Information Technology. Lou does not seem to have any scandal, his appearance on GFW's rulebook may just mean he is personally closed to the GFW operation.</i>		
马馼	prefix.standard	1

...continued from previous page		
Term	GFW Rule	Pages Affected
馬馱	prefix.standard zh.wikipedia.org/zh-cn/\$term zh.wikipedia.org/zh-hans/\$term	1
<i>Ma Wen, a female Party official handled the Bo Xilai case.</i>		
钱其琛 <i>Qian Qichen, Foreign Minister from 1988 to 1998.</i>	zh.wikipedia.org/zh-tw/\$term	2 钱其琛 钱其琛_(国民党)
宋任穷 宋任窮 <i>Song Renqiong.</i>	broad.standard broad.standard	2 Talk:宋任穷 1
陶驷驹	prefix.standard zh.wikipedia.org/zh/\$term zh.wikipedia.org/zh-cn/\$term zh.wikipedia.org/zh-tw/\$term	1
陶駟駒	prefix.standard zh.wikipedia.org/zh/\$term zh.wikipedia.org/zh-cn/\$term zh.wikipedia.org/zh-hk/\$term	1
<i>Tao Siju, Minister of Public Security from 1990 to 1998.</i>		
庹震	prefix.standard zh.wikipedia.org/zh/\$term zh.wikipedia.org/zh-hk/\$term zh.wikipedia.org/zh-sg/\$term	1
<i>Tuo Zhen, Propaganda Chief of Guangdong Province, responsible for the Southern Weekly Incident in Jan 2013.</i>		
王立军	prefix.standard	4 王立军事件 王立军徇私枉法、叛逃、滥用职权、受贿案 王立军案
<i>Wang Lijun, ex-Police Chief of Chongqing, right-hand man of Bo Xilai. His defection to the US Consulate in Feb 2012 started the biggest political turmoil in recent years, which led to the downfall of Bo Xilai.</i>		
张立昌 <i>Zhang Lichang, infamous ex-Party Chief of Tianjin.</i>	prefix.standard	1
周强 <i>Zhou Qiang.</i>	prefix.standard	1
Miscellaneous		
方舟子 <i>Fang Zhouzi, an internet celebrity, famous for picking fights with other celebrities, usually in the name of fighting fraud.</i>	broad.standard	4 Talk:方舟子 User_talk:我是方舟子 Category:方舟子支持者
宋彬彬 <i>Song Binbin, daughter of Song Renqiong. She is widely regarded as responsible for beating Bian Zhongyun to death in the Cultural Revolution.</i>	broad.standard	3 Talk:宋彬彬 User_talk:宋彬彬
王小丫 <i>Wang Xiaoya, a TV host, wife of Cao Jianming. Rumors say she was involved in the corruption case of Wang Yi.</i>	zh.wikipedia.org/zh-tw/\$term zh.wikipedia.org/zh-sg/\$term	1 王小丫
王姓 <i>Surname Wang. This page is about the family name Wang.</i>	prefix.standard	1
俞丽萍 <i>Yu Liping (Jennifer Yu), president of Rothschild's Greater China Region, whose husband is rumored to be Jiang Zemin's adopted son.</i>	prefix.standard	1
张钰	broad.standard	6 张钰_(演员) 张钰哲 张钰钟 Talk:张钰_(演员) User_talk:张钰涵
張鈺	broad.standard	7 張鈺_(乒乓球運動員) 張鈺_(消歧義) 張鈺_(運動員) Talk:張鈺_(乒乓球運動員) Talk:張鈺_(運動員) User_talk:張鈺燁
<i>Zhang Yu, likely this is due to the actress who publicized sexual video with movie directors in Nov 2006.</i>		

Table 8.4: GFW Rulebook: People, dissidents, officials & miscellaneous

8.5 Regional

This section covers GFW rules regarding Hong Kong, Taiwan, Tibet and Uyghur. The numbers suggest China censors worry about Hong Kong much more than Taiwan.

- **Hong Kong**, 52 rules total. 17 rules are for Hong Kong politicians in the pan-democratic camp, 21 rules are for protesting rallies in Hong Kong.
- **Taiwan**, 15 rules total. 6 rules are for presidential candidates in the 2012 election, 4 are about presidential elections in Republic of China (a.k.a. Taiwan).
- **Tibet**, 39 rules total. 5 rules target the English Wikipedia.
- **Uyghur**, 22 rules total. 2 rules target the English Wikipedia.

In addition, on GFW's rulebook for HTTP response filtering (see Section 5 and Table 5.1), there are five terms targeting Wikipedia that are related to Tibet. The term “达赖喇嘛” (Dalai Lama) is one of them and it interrupts many sensitive and non-sensitive Wikipedia pages.

Term	GFW Rule	Pages Affected
Regional: Hong Kong		
司徒华	broad.standard	2 玛莎·司徒华
司徒華	broad.standard	4 司徒華特島 司徒華獎 Talk:司徒華
Szeto Wah	prefix.standard	1 Szeto Wah
<i>Szeto Wah, (1931 - 2011/01/02), Hong Kong politician and activist. Iconic figure of Hong Kong democracy movement. Founder of 支联会 (the Hong Kong Alliance in Support of Patriotic Democratic Movements in China).</i>		
李柱铭	broad.standard	1
李柱銘	broad.standard	4 Talk:李柱銘 User:香港國父李柱銘 User_talk:香港國父李柱銘
<i>Martin Lee Chu-Ming, prominent Hong Kong politician, regarded as "Father of Hong Kong Democracy" by the pan-democracy camp.</i>		
李卓人	broad.standard	2 Talk:李卓人
<i>Lee Cheuk-Yan, Hong Kong politician and activist. Chairman of the Hong Kong Alliance in Support of Patriotic Democratic Movements in China.</i>		
何俊仁	broad.standard	2 Talk:何俊仁
<i>Albert Ho Chun-Yan, member of the Legislative Council, former chairman of the Democratic Party.</i>		
梁国雄	broad.standard	1
梁國雄	broad.standard	4 Talk:梁國雄 User:曾鈺成：梁國雄議員，請你冷靜！ User_talk:曾鈺成：梁國雄議員，請你冷靜！
<i>Leung Kwok-Hung, Hong Kong activist, nickname "longhair", member of the Legislative Council.</i>		
刘慧卿	broad.standard	1
劉慧卿	broad.standard	2 Talk:劉慧卿
<i>Emily Lau Wai-Hing, Hong Kong politician, member of the Legislative Council, current chairwoman of Democratic Party.</i>		
张文光	broad.standard	1 张文光_(清末民初)
張文光	broad.standard	3 張文光_(崇禎進士) Talk:張文光
<i>Cheung Man-Kwong, Hong Kong politician, former member of the Legislative Council, member of the pan-democracy camp.</i>		
劉迺强	prefix.standard	1
<i>Lau Nai-Keung, pro-CCP Hong Kong politician.</i>		
霍英東	prefix.standard	2 霍英東家族
<i>Henry Fok Ying-Tung, (1923 - 2006/10/28), Hong Kong businessman, regarded by many as the most powerful Hongkonger in the politics of China.</i>		
衛奕信	prefix.standard	2 衛奕信徑
<i>David Clive Wilson, Baron Wilson of Tillyorn, 27th Governor of Hong Kong (from 1987 to 1992).</i>		
香港游行	broad.standard	0
香港遊行	broad.standard	1 Category:香港遊行
<i>Hong Kong march.</i>		
七一游行	prefix.standard	1

continued on next page...

...continued from previous page

Term	GFW Rule	Pages Affected
七一遊行	prefix.standard zh.wikipedia.org/zh-tw/\$term zh.wikipedia.org/zh-hant/\$term	1
七一大遊行	zh.wikipedia.org/zh-tw/\$term zh.wikipedia.org/zh-hk/\$term zh.wikipedia.org/zh-hant/\$term	1 七一大遊行
香港七一游行	zh.wikipedia.org/zh-hant/\$term	1 香港七一游行
香港七一遊行	zh.wikipedia.org/zh-tw/\$term zh.wikipedia.org/zh-hk/\$term zh.wikipedia.org/zh-hant/\$term	1 香港七一遊行
香港七一大遊行	prefix.standard	1
<i>various terms for the Hong Kong 1 July Marches, an annual protest rally with hundreds of thousands of participants.</i>		
2003年香港七一游行	prefix.standard	1
<i>Hong Kong 1 July March in 2003.</i>		
七一效应	broad.standard	2 Talk:七一效应
七一效應	broad.standard	2 Talk:七一效應
<i>July 1st effect, a term describing the effect of pro-democracy sentiment on voters voting for candidates in the pan-democracy camp.</i>		
2007年反对人大否决普选游行	broad.standard	1
2007年反對人大否决普選遊行	broad.standard	2 Talk:2007年反對人大否决普選遊行
<i>2007 rally in protesting the veto of universal suffrage by China.</i>		
2007年區議會選舉	prefix.standard	1
2007年香港區議會選舉	prefix.standard	5 2007年香港區議會選舉候選人名單 2007年香港區議會選舉參選名單 2007年香港區議會選舉結果 ...
<i>2007 district legislature election.</i>		
2008年争取2012双普选大游行	broad.standard	1
2008年爭取2012雙普選大遊行	broad.standard	2 Talk:2008年爭取2012雙普選大遊行
<i>2008 rally for double universal suffrage in 2012. "double" refers to the elections for the chief executive and legislature.</i>		
2012年双普选	broad.standard	1
2012年雙普選	broad.standard	3 Talk:2012年雙普選 Category:支持香港2012年雙普選的維基人
<i>double universal suffrage in 2012. "double" refers to the elections for the chief executive and legislature.</i>		
民间人权阵线	broad.standard	0
民間人權陣線	broad.standard	2 Talk:民間人權陣線
<i>Civil Human Rights Front, a platform affiliating almost all the pan-democracy camps, organizer of the annual July 1st Marches.</i>		
民主党_(香港)	broad.standard	0
民主黨_(香港)	broad.standard	5 Talk:民主黨_(香港) Template:民主黨_(香港)/meta/color Template:民主黨_(香港)/meta/shortname Category:民主黨_(香港)
<i>Democratic Party (Hong Kong).</i>		
四五行动	broad.standard	1
四五行動	broad.standard	2 Category:四五行動成員
<i>April Fifth Action, a socialist organization, but aligning more with the pan-democracy camp and opposing CCP regime.</i>		
香港民主	broad.standard	23 香港民主党 香港民主派 香港民主運動發展史 香港民主黨 Talk:香港民主民生協進會 Template:User_香港民主黨 ...
<i>Hong Kong democracy.</i>		
香港独立主义	broad.standard	1
香港獨立主義	broad.standard	1
<i>Hong Kong independentism.</i>		
香港泛民主派	broad.standard	13 香港泛民主派人士列表 香港泛民主派組織 Talk:香港泛民主派 Template:香港泛民主派控制的區議會選區 ...

continued on next page...

...continued from previous page		
Term	GFW Rule	Pages Affected
<i>Hong Kong pan-democracy camp.</i>		
香港市民支援愛國民主運動聯合會	zh.wikipedia & \$term	2 Talk:香港市民支援愛國民主運動聯合會
<i>Hong Kong Alliance in Support of Patriotic Democratic Movements in China.</i>		
Regional: Taiwan		
馬英九	prefix.standard	3 馬英九政府 馬英九總統任期
<i>Ma Ying-Jeou, President of the Republic of China (a.k.a. Taiwan). Ma was the (winning) presidential candidate of Kuomintang in the 2012 election.</i>		
吳敦義	prefix.standard	2 吳敦義內閣
<i>Wu Den-Yih, Vice President of the Republic of China (a.k.a. Taiwan), running mate of Ma Ying-Jeou in the 2012 election.</i>		
蔡英文	prefix.standard	2 蔡英文_(消歧義)
<i>Tsai Ing-Wen, former chairwoman of the Democratic and Progressive Party (DPP). Tsai was the presidential candidate of DPP in the 2012 election.</i>		
蘇嘉全	prefix.standard	1
<i>Su Chia-Chyuan, running mate of Tsai Ing-Wen in the 2012 election.</i>		
宋楚瑜	prefix.standard	1
<i>James Soong Chu-Yu, founder and chairman of the People First Party (PFP). Soong was the presidential candidate of PFP in the 2012 election.</i>		
林瑞雄	prefix.standard	1
<i>Lin Ruey-Shiung, running mate of Soong Chu-Yu in the 2012 election.</i>		
洪哲胜	broad.standard	0
洪哲勝	broad.standard	2 Talk:洪哲勝
<i>Cary S. Hung, Taiwan activist, with involvement in mainland China's democracy movement.</i>		
蒙藏委员会	broad.standard	1
蒙藏委員會	broad.standard	7 中華民國蒙藏委員會 Talk:中華民國蒙藏委員會 Talk:蒙藏委員會 Category:中華民國蒙藏委員會 ...
<i>Mongolian and Tibetan Affairs Commission, a government branch of Republic of China (a.k.a. Taiwan).</i>		
台灣選舉	prefix.standard	2 台灣選舉地理
<i>Taiwan election.</i>		
台灣總統	prefix.standard	3 台灣總統府 台灣總統直選列表
<i>Taiwan president.</i>		
中華民國總統	prefix.standard	18 中華民國總統列表 中華民國總統及副總統之公民直接選舉與罷免 中華民國總統官邸 中華民國總統府 ...
<i>president of Republic of China.</i>		
2012年中华民国总统选举	prefix.standard	2 2012年中华民国总统选举民意调查
<i>2012 presidential election in Republic of China (a.k.a. Taiwan).</i>		
台灣獨立建國聯盟	prefix.standard	1
<i>World United Formosans for Independence.</i>		
Regional: Tibet		
Dalai.Lama	en.wikipedia.org/wiki/\$term	0-EN:11 Dalai.Lama(en) Dalai.Lamas(en) Dalai.Lama.V(en) Dalai.Lama.(song)(en) Dalai.Lama.Renaissance(en) ...
达赖喇嘛	prefix.standard	3 达赖喇嘛·丹增嘉措 达赖喇嘛文化复兴
<i>Dalai Lama.</i>		
Tenzin.Gyatso	en.wikipedia.org/wiki/\$term	0-EN:4 Tenzin.Gyatso(en) Tenzin.Gyatso_(Dalai.Lama)(en) Tenzin.Gyatso,_14th.Dalai.Lama(en) ...
<i>Religious name of the 14th Dalai Lama.</i>		
十四世达赖	broad.standard	6 第十四世达赖喇嘛 第十四世达赖喇嘛·丹增嘉措 Talk:第十四世达赖喇嘛 Talk:第十四世达赖喇嘛·丹增嘉措 ...
十四世達賴	broad.standard	4 第十四世達賴喇嘛 第十四世達賴喇嘛·丹增嘉措 第十四世達賴喇嘛丹增嘉措 Talk:第十四世達賴喇嘛
<i>14th Dalai (Lama).</i>		
达赖集团	broad.standard	1
達賴集團	broad.standard	2 Talk:達賴集團

...continued from previous page		
Term	GFW Rule	Pages Affected
<i>Dalai Group.</i>		
平措尼珍	broad.standard	1
<i>Phuntsog Nyidron, a Tibetan Buddhist nun, jailed from 1989 to 2004.</i>		
當知項欠	zh.wikipedia.org/zh-cn/\$term	1 當知項欠
<i>Dhondup Wangchen, producer of the documentary Jigdlrel (不再恐惧, Leaving Fear behind), in prison since Mar 2008. The co-producer 久美嘉措 (Jigme Gyatso) is not on GFW's rulebook.</i>		
夏格巴·旺秋德丹	broad.standard	1
<i>Tsepon Wangchuk Deden Shakabpa, (1907 - 1989/02/23), Tibetan historian.</i>		
西藏问题	broad.standard	2 Talk:西藏问题
西藏問題	broad.standard	2 聯合國關於西藏問題的決議
<i>Tibet issues.</i>		
Flag_of_Tibet	en.wikipedia.org/wiki/\$term	0-EN:2 Flag_of_Tibet(en) Flag_of_tibet(en)
雪山狮子旗	broad.standard	1
雪山獅子旗	broad.standard	2 Talk:雪山獅子旗
<i>Snow Mountain and Lion Flag, the flag of Tibet.</i>		
西藏流亡政府	broad.standard	12 其美仁增_(西藏流亡政府) 西藏流亡政府历届噶伦赤巴列表 Template:Country_data_西藏流亡政府 ...
<i>Tibetan government in exile.</i>		
达兰萨拉	broad.standard	2 Talk:达兰萨拉
達蘭薩拉	broad.standard	1
<i>Dharmasāla, a town in India, home to Tibetan refugees and the Tibetan government in exile.</i>		
藏独	prefix.standard	1
藏獨	broad.standard	6 藏獨運動 西藏獨立運動 Template:User_反藏獨 Template:User_藏獨 Category:反對藏獨的維基人 ...
<i>Tibetan independence.</i>		
Tibetan_Independence_Movement	en.wikipedia.org/wiki/\$term	0-EN:2 Tibetan_Independence_Movement(en) Tibetan_independence_movement(en)
西藏独立运动	broad.standard	2 Talk:西藏独立运动
<i>Tibetan independence movement.</i>		
Students_for_a_Free_Tibet	en.wikipedia.org/wiki/\$term	0-EN:1 Students_for_a_Free_Tibet(en)
自由西藏学生运动	broad.standard	2 自由西藏学生运动组织
自由西藏學生運動	broad.standard	1
<i>Students for a Free Tibet Movement.</i>		
国际声援西藏运动	prefix.standard	1
<i>zh.wikipedia.org/zh-sg/\$term</i>		
<i>International Campaign for Tibet.</i>		
西藏抗暴紀念日	broad.standard	1
西藏人民起义日	broad.standard	2 Talk:西藏人民起义日
西藏人民起義日	broad.standard	1
<i>Tibetan Uprising Day, observed on March 10 to commemorate the 1959 Tibetan uprising.</i>		
囊帕拉枪杀事件	broad.standard	2 Talk:囊帕拉枪杀事件
囊帕拉槍殺事件	broad.standard	2 Talk:囊帕拉槍殺事件
<i>Nangpa La Shooting Incident. On September 30, 2006, a group of Tibetan refugees were shot by Chinese border guards when passing Nangpa La.</i>		
2008年藏人抗议事件	broad.standard	1
2008年藏人抗議事件	broad.standard	1
<i>Tibetan Protests in 2008.</i>		
2012年藏区骚乱	prefix.standard	1
<i>Riots in Tibetan Region in 2012.</i>		
藏區連環自焚事件	prefix.standard	1
<i>Serial Self-immolations in Tibetan Region, more than 120 cases from 2009 to 2013.</i>		

...continued from previous page		
Term	GFW Rule	Pages Affected
西藏生死書	prefix.standard	1
<i>The Tibetan Book of Living and Dying, a book on Tibetan spirituality, not politically sensitive.</i>		
写给西藏的歌	broad.standard	0
寫給西藏的歌	broad.standard	1
<i>Songs for Tibet - The Art Of Peace, an album by western musicians, released in 2008.</i>		
Regional: Uyghur		
East_Turkestan_Independe nce_Movement	en.wikipedia.org/wiki/\$term	0-EN:3 East_Turkestan_Independence_Movement(en) East_Turkestan_independence_movement(en) ...
Rebiya_Kadeer	en.wikipedia.org/wiki/\$term	0-EN:1 Rebiya_Kadeer(en)
热比亚	prefix.standard	2 热比亚·卡德尔
热比娅	broad.standard	2 热比娅·卡德尔 Talk:热比娅·卡德尔
熱比婭	broad.standard	2 熱比婭·卡德爾
<i>Rebiya Kadeer, Uyghur activist.</i>		
侯赛因江	broad.standard	2 Talk:侯赛因江
侯賽因江	broad.standard	0
<i>Höseyinjan Jélil, East Turkestan independence activist, sentenced to life in prison in 2007.</i>		
买买提明·艾孜来提	broad.standard	2 Talk:买买提明·艾孜来提
買買提明·艾孜來提	broad.standard	0
<i>Mehmet Emin Hazret, East Turkestan independence activist.</i>		
畏吾兒	prefix.standard	1
<i>Uyghur, an alternative term for the current standard term “维吾尔”, which is not on GFW's rulebook.</i>		
東突	broad.standard	18 後東突厥 東突厥 東突厥伊斯蘭共和國 東突厥可汗列表 東突厥戰爭 東突厥斯坦 東突厥斯坦伊斯蘭運動 ...
<i>short name for East Turkestan.</i>		
东突厥斯坦	broad.standard	15 东突厥斯坦人民共和国 东突厥斯坦伊斯兰国 东突厥斯坦伊斯兰运动 东突厥斯坦信息中心 东突厥斯坦共和国 ...
<i>East Turkestan.</i>		
东土耳其斯坦解放组织	broad.standard	1
東土耳其斯坦解放組織	broad.standard	1
<i>East Turkestan Liberation Organization.</i>		
东土耳其斯坦信息中心	broad.standard	2 Talk:东土耳其斯坦信息中心
<i>East Turkestan Information Center.</i>		
世界维吾尔青年代表大会	broad.standard	2 Talk:世界维吾尔青年代表大会
世界維吾爾青年代表大會	broad.standard	1
<i>World Uyghur Youth Congress. It merged with a few other groups and formed 世界维吾尔代表大会 (World Uyghur Congress), which is not on GFW's rulebook.</i>		
新疆独立运动	broad.standard	2 Talk:新疆独立运动
新疆獨立運動	broad.standard	1
<i>Xin Jiang independence movement.</i>		
乌鲁木齐七·五事件	zh.wikipedia.org/zh-hans/\$term	1 乌鲁木齐七·五事件
<i>Urumuqi July 5th Incident, a riot in 2009.</i>		
烏魯木齊武警毆打香港 記者事件	prefix.standard	1
<i>an incident on Sep 4, 2009 in which the Urumuqi armed police beat three Hong Kong reporters.</i>		
愛的十個條件	zh.wikipedia.org/zh-hant/\$term	1 愛的十個條件
<i>The 10 Conditions of Love, a documentary about Rebiya Kadeer.</i>		

Table 8.5: GFW Rulebook: Regional: Hong Kong, Taiwan, Tibet, Uyghur

8.6 Falungong

Falungong, a religious group persecuted in China, used to be the motivation and focus of GFW. There are 18 rules here, none seem to be new. There are many more Wikipedia articles related to Falungong but they are not on GFW's rulebook for HTTP request filtering. The reason is that GFW has five Falungong terms in its rulebook for HTTP response filtering, i.e. “flg”, “falun”, “法轮功”, “法轮大法 & 明慧网”, and “新唐人电视台 & 神韵艺术团”. These rules effectively block all Falungong-related Wikipedia articles, plus many more sensitive or non-sensitive Wikipedia pages not directly related to Falungong.

Term	GFW Rule	Pages Affected
法轮	broad.standard	42 法轮功 法轮_(消歧义) 中央处理法轮功问题领导小组办公室 护国法轮寺 法轮_(佛教) 法轮功在各国的发展及影响 ...
法輪	broad.standard	44 法輪功 三轉法輪 法輪_(高曇晟) 法輪功受迫害真相聯合調查團 法輪功在台灣 法輪功在香港 法輪功歷史 ...
<i>Falun, dharma wheel, the first two characters of Falungong.</i>		
李洪志	broad.standard	3 李瑞_(李洪志妻子) Talk:李洪志
<i>Li Hongzhi, founder of Falungong.</i>		
退党潮	broad.standard	2 Talk:退党潮
退黨潮	broad.standard	1
<i>wave of resignations from the Chinese Communist Party.</i>		
九评共产党	broad.standard	2 Talk:九评共产党
九評共產黨	broad.standard	1
<i>Nine Commentaries on CCP, a series of articles published by Falungong organization.</i>		
苏家屯事件	broad.standard	1
蘇家屯事件	broad.standard	2 Talk:蘇家屯事件
<i>Sujiatun incident, an alleged concentration camp in Sujiatun in which thousands of Falungong practitioners were allegedly tortured.</i>		
大纪元时报	broad.standard	3 Talk:大纪元时报 Talk:大纪元时报/存档1
大紀元時報	broad.standard	1
<i>Epoch Times, a newspaper operated by the Falungong organization.</i>		
新唐人电视台	broad.standard	2 Talk:新唐人电视台
新唐人電視台	broad.standard	2 Talk:新唐人電視台
<i>New Tang Dynasty Television, a TV station operated by the Falungong organization.</i>		
人民报_(中文)	broad.standard	2 Talk:人民报_(中文)
人民報_(中文)	broad.standard	0
<i>People's Newspaper, renminbao.com, a news site operated by the Falungong organization.</i>		
天安门自焚事件	broad.standard	2 Talk:天安门自焚事件
天安門自焚事件	broad.standard	0
<i>Tiananmen self immolation incident. A self immolation incident in the Tiananmen Square on Jan 23 2001. China authority claims the person is a Falungong practitioner, while the Falungong organization claims it was staged to defame Falungong.</i>		
希望之声国际广播电台	broad.standard	2 Talk:希望之声国际广播电台
<i>Sound of Hope Radio Network, a radio station affiliated with Falungong.</i>		

Table 8.6: GFW Rulebook: Falungong

8.7 Tiananmen

Among all sensitive topics in China, the most sensitive is the Tiananmen Square Protest and Crackdown in 1989, also known as the June 4th Incident, June 4th Movement, the June 4th Massacre, etc.

China authority has been very successful in removing the Tiananmen memory from Chinese people. There is nearly zero mentioning of this historical event inside China, including news, books, TV, and internet. As a result, most people in China do not know what happened in 1989, and the majority of the younger generation have not even heard of it. GFW has been instrumental in constructing this memory hole.

This section alone contains 116 such rules. In addition to that, we will also see many obscure terms motivated by Tiananmen Incident in Section 8.9. However, one surprising finding is that there is no Tiananmen related rules in GFW's HTTP response filtering (see Section 5), hence many Tiananmen related articles are actually accessible in China, e.g. <http://en.wikipedia.org/wiki/Tankman>. Also, terms blocked by prefix rules can be accessed by a non-blocked variant. For example, “64事件” (June 4th Incident) is blocked by prefix rules only for the “/wiki/”, “/zh-cn/” and “/zh-hk/” variants, the other variants are all accessible in China, e.g. <http://zh.wikipedia.org/zh/64事件>.

This is exactly the opposite to the Falungong case. For Falungong, there are only 18 rules in GFW's HTTP request scan, but the five rules in GFW's HTTP response scan effectively block all Falungong-related content.

We roughly divide the terms into the following groups:

- Those targeting the English Wikipedia. 5 total.
- June 4th event.
- Memorial events.
- People's Liberation Army units which participated in the June 4th crackdown.
- Publications and songs.

The grouping is not clearcut here. Common prefix is another consideration for grouping, so that terms with common prefix are together, even though they might have different meanings. Also note that one Wikipedia URL might offend multiple GFW rules, because China authority really had thrown many many rules into GFW.

Term	GFW Rule	Pages Affected
Terms Targeting English Wikipedia		
Tiananmen_Square_Protests_of_1989	en.wikipedia.org/wiki/\$term prefix.standard	0-EN:3 Tiananmen_Square_Protests_of_1989(en) Tiananmen_Square_Protests_of_1989(en) Tiananmen_square_protests_of_1989(en)
<i>Note that GFW has this term for both English and Chinese Wikipedias, even though there is no such page in the Chinese version.</i>		
Tank_Man	en.wikipedia.org/wiki/\$term prefix.standard	0-EN:3 Tank_Man(en) Tank_man(en) Tank_Man_(Battle_Angel_Alita)(en)
<i>Note that GFW has this term for both English and Chinese Wikipedias, even though there is no such page in the Chinese version. However, the article en.wikipedia.org/wiki/Tankman is accessible.</i>		
Tiananmen_Massacre	en.wikipedia.org/wiki/\$term	0-EN:2 Tiananmen_Massacre(en) Tiananmen_massacre(en)
Tiananmen_Papers	en.wikipedia.org/wiki/\$term	0-EN:1 Tiananmen_Papers(en)
<i>A book published in 2001. The Chinese version is published under the name 中国六四真相 (The Truth of June 4th in China) the same year.</i>		
20th_anniversary_Tiananmen_square_incident_march	wiki/\$term	0-EN:1 20th_anniversary_Tiananmen_square_incident_march(en)
<i>Note that the rule's Wikipedia part is simply “wiki/”, this is due to the length of the rule because GFW rules are limited to 64 bytes in size.</i>		
June 4th Event		
June_4th	prefix.standard	0
<i>Note that this rule is for the Chinese Wikipedia. This article was deleted in Sep 2013, see the notice at June_4th.</i>		
Liusishijian	prefix.standard zh.wikipedia.org/zh-hans/\$term	1 Liusishijian
<i>This is the pinyin for 六四事件 (June 4th Incident).</i>		

continued on next page...

...continued from previous page		
Term	GFW Rule	Pages Affected
1989年	broad.standard	232 1989年天安門事件 1989年电影 河_(1989年) 香港1989年 Talk:1989年 Category:1989年 Category_talk:1989年足球 ...
<i>Year 1989. This rule affects 232 Wikipedia pages.</i>		
1989天安门	prefix.standard zh.wikipedia.org/zh-cn/\$term	1
1989天安門 1989 Tiananmen.	zh.wikipedia.org/zh-cn/\$term	1 1989天安門
6.4事件	prefix.standard zh.wikipedia.org/zh/\$term zh.wikipedia.org/zh-cn/\$term zh.wikipedia.org/zh-hant/\$term	1
64事件	prefix.standard zh.wikipedia.org/zh-cn/\$term zh.wikipedia.org/zh-hk/\$term	1
<i>June 4th Incident.</i>		
8964	prefix.standard	1
<i>the year, month and date of the June 4th crackdown.</i>		
89学潮	prefix.standard	1
89学运	prefix.standard	1
<i>89 student movement.</i>		
89运动	prefix.standard zh.wikipedia.org/zh-cn/\$term	1
<i>89 movement.</i>		
89民运	prefix.standard zh.wikipedia.org/zh-cn/\$term zh.wikipedia.org/zh-hans/\$term	1
<i>89 democracy movement.</i>		
八九	prefix.standard	14 八九动乱 八九天安門事件 八九学生民主运动 八九学运 八九學潮 八九式擲彈筒 八九式機槍 八九民主運動 ...
<i>eight nine.</i>		
六四	prefix.standard zh.wikipedia.org/zh/\$term zh.wikipedia.org/zh-tw/\$term zh.wikipedia.org/zh-hk/\$term zh.wikipedia.org/zh-hant/\$term	44 六四風波 六四集會 六四人物 六四大屠殺 六四紀念館 六四綠卡 六四血卡 六四式手枪 六四式輕戰車 ...
<i>six four, the most common short name for the Tiananmen Incident. Note that many terms with this prefix are blocked by the broad rule, see below.</i>		
六四事件	broad.standard zh-yue.wikipedia.org/wiki/\$term	34 六四事件人物 六四事件反应 六四事件纪念活动 Talk:六四事件 Template:六四事件 Category:六四事件 ...
<i>June 4th Incident.</i>		
六四歌曲	broad.standard	4 Talk:六四歌曲 User:飞剑鱼/六四歌曲 Category:六四歌曲
<i>June 4th songs.</i>		
六四诗集	broad.standard	2 Talk:六四诗集
六四詩集	broad.standard	0
<i>June 4th Poetry Collection.</i>		
六四襟章	broad.standard	2 Talk:六四襟章
<i>June 4th badge.</i>		
六四内部日記	broad.standard	1
<i>June 4th Internal Diary, a book published in 2006 based on diaries of 陆超祺 (Lu Chaoqi), then-acting chief editor of People Daily.</i>		
六四18週年	broad.standard	1
<i>18th anniversary of June 4th Incident.</i>		

...continued from previous page		
Term	GFW Rule	Pages Affected
陆肆事件	wikipedia.org & \$term	1
陸肆事件	wikipedia.org & \$term	1 User:紀念陸肆事件貳拾壹周年
<i>June 4th Incident.</i>		
春夏之交的政治風波	prefix.standard	1
<i>“political disturbance between spring and summer”, an euphemistic term used by China authority in the 1990’s to describe the June 4th Incident.</i>		
北京民運	prefix.standard	1
	zh.wikipedia.org/zh/\$term	
	zh.wikipedia.org/zh-tw/\$term	
<i>Beijing democracy movement, this article redirects to the June 4th Incident article.</i>		
北京高校学生自治联合会	broad.standard	1
北京高校學生自治聯合會	broad.standard	0
<i>Beijing Students Autonomous Federation, which is a major organizer of the Tiananmen protest.</i>		
全美中国学生学者自治联合会	zh.wikipedia.org/zh-tw/\$term	1 全美中国学生学者自治联合会
<i>Independent Federation of Chinese Students and Scholars of USA, an organization formed in 1989.</i>		
学自联	broad.standard	0
學自聯	broad.standard	0
<i>Short name for the above organization. Wikipedia does not have articles with these two titles.</i>		
民主女神	broad.standard	3 民主女神像 Talk:民主女神
<i>Goddess of Democracy, a statue created in the Tiananmen Square in 1989.</i>		
民主之神宣言	prefix.standard	1
	zh.wikipedia.org/zh-tw/\$term	
<i>a manifesto published during the establishment of the Goddess of Democracy.</i>		
天安门	prefix.standard	19 天安门 _(消歧义) 天安门 _(电影) 天安门 _(纪录片) 天安门 东 站 天安门四君子 天安门国旗护卫队 天安门城楼 ...
天安門	zh.wikipedia.org/zh-cn/\$term	
	zh.wikipedia.org/zh-tw/\$term	
11 天安門 天安門 _(紀錄片) 天安門事件 天安門城樓 天安門大屠殺 天安門屠殺 天安門廣場 天安門廣場 天安門自焚		
<i>Tiananmen. Note that many terms with this prefix are blocked by the broad rule, see below.</i>		
天安門母親	prefix.standard	1
	zh.wikipedia.org/zh-cn/\$term	
<i>Tiananmen Mothers, an activist group of families of June 4th victims.</i>		
天安门母亲运动	broad.standard	1
天安門母親運動	broad.standard	0
<i>Tiananmen Mothers Movement.</i>		
天安门事件	broad.standard	6 八九天安门事件 六四天安门事件 四五天安门事件 天安门事件 _(文革时期) Talk:天安门事件
天安門事件	broad.standard	16 1989年天安門事件 六四天安門事件 四五天安門事件 User:天安門事件 User:鄧小平是六四天安門事件的壞蛋！ ...
	zh-yue.wikipedia.org/wiki/\$term	
<i>Tiananmen Incident. The traditional Chinese version is one of four terms that target the Yue (Cantonese) Wikipedia.</i>		
天安門屠殺	prefix.standard	1
	zh.wikipedia.org/zh-cn/\$term	
	zh.wikipedia.org/zh-hk/\$term	
<i>Tiananmen massacre.</i>		
天安门(纪录片)	broad.standard	2 Talk:天安门(纪录片)
<i>Tiananmen (documentary), a PBS documentary on the Tiananmen Incident. The English title is “The Gate of Heavenly Peace”.</i>		
天安门文件	broad.standard	1
天安門文件	broad.standard	2 Talk:天安門文件
<i>Tiananmen Papers, a book published in 2001. Its Chinese title is 中国六四真相(The Truth of June 4th in China).</i>		
中國六四真相	prefix.standard	1

...continued from previous page		
Term	GFW Rule	Pages Affected
<i>The Truth of June 4th in China, Chinese title of the book Tiananmen Papers.</i>		
承天门	prefix.standard	1
<i>Chengtianmen, the original name of Tiananmen, renamed to Tiananmen in 1651. An obscure term.</i>		
建国门	prefix.standard	14 建国门_(北京) 建国门_(西安) 建国门事件 建国门外 建国门桥 建国门站 建国门街道 建国门外大街 ...
<i>Jianguomen, a gate and area in Beijing. It is blocked for two possible reasons, one is killings in this area in 1989, the other is a mass shooting incident on Oct 20, 1994.</i>		
王維林	broad.standard	3 Talk:王維林 User.talk:王維林
王維林	broad.standard	1
<i>Wang Weilin, a widely circulated name of the Tank Man, whose identity and whereabouts remain a mystery.</i>		
黄雀行动	broad.standard	1
黄雀行動	broad.standard	1
<i>Operation Siskin, or Operation Yellow Bird, an operation to rescue Chinese students and citizens who were wanted by the China authority after the June 4th crackdown.</i>		
全球華人大遊行	zh.wikipedia.org/zh-tw/\$term zh.wikipedia.org/zh-hk/\$term	1 全球華人大遊行
<i>Worldwide Chinese Rallies, which happened overseas on May 28, 1989, to support the protests in China.</i>		
Memorial Events		
維多利亞公園	prefix.standard	9 維多利亞公園_(消歧義) 維多利亞公園六四燭光晚會 維多利亞公園年宵市場 維多利亞公園站 維多利亞公園鎮 ...
<i>Victoria Park, a public park in Hong Kong. It is the venue of the annual vigil for the June 4th Incident.</i>		
维园六四烛光晚会	broad.standard	0
維園六四燭光晚會	broad.standard	3 香港維園六四燭光晚會 Talk:維園六四燭光晚會
<i>Victoria Park June 4th Candlelight Vigil.</i>		
維園燭光晚會	zh.wikipedia.org/zh-cn/\$term zh.wikipedia.org/zh-hant/\$term	1 維園燭光晚會
<i>Victoria Park Candlelight Vigil.</i>		
香港六四集會	prefix.standard	1
<i>Hong Kong June 4th Assembly.</i>		
香港六四燭光晚會	prefix.standard zh.wikipedia.org/zh-tw/\$term zh.wikipedia.org/zh-hk/\$term	1
<i>Hong Kong June 4th Candlelight Vigil.</i>		
People's Liberation Army (PLA) Units		
中国人民解放军第20集团军	prefix.standard	1
<i>the 20th group army of PLA.</i>		
中国人民解放军第27集团军	prefix.standard	1
<i>the 27th group army of PLA.</i>		
38軍	prefix.standard	1
第38军	prefix.standard	2 第38军_(日本陆军)
中国人民解放军第38集团军	prefix.standard	1
<i>zh.wikipedia.org/zh-cn/\$term</i>		
中國人民解放軍第38集團軍	prefix.standard	1
<i>various terms for the 38th group army of PLA, which committed the most killings in the June 4th crackdown.</i>		
39军	prefix.standard	1
中国人民解放军第39集团军	prefix.standard	1

...continued from previous page		
Term	GFW Rule	Pages Affected
<i>short name and full name of the 39th group army of PLA.</i>		
中国人民解放军第54集团军	zh.wikipedia.org/zh-cn/\$term	1 中国人民解放军第54集团军
<i>the 54th group army of PLA.</i>		
65军	prefix.standard	1
	zh.wikipedia.org/zh-hant/\$term	
66455部队	prefix.standard	1
中国人民解放军第65集团军	zh.wikipedia.org/zh-cn/\$term	1 中国人民解放军第65集团军
<i>short name, internal code, and full name of the 65th group army of PLA.</i>		
Publications & Songs		
民运歌曲	prefix.standard	1
	zh.wikipedia.org/zh/\$term	
	zh.wikipedia.org/zh-tw/\$term	
	zh.wikipedia.org/zh-hk/\$term	
	zh.wikipedia.org/zh-hant/\$term	
民運歌曲	prefix.standard	0
<i>Songs of democracy movement. This page redirects to 六四歌曲 (June 4th songs.)</i>		
民主歌聲獻中華	prefix.standard	1
	zh.wikipedia.org/zh-hk/\$term	
	zh.wikipedia.org/zh-hant/\$term	
<i>Concert for Democracy In China, a large concert on May 27, 1989 in Kong Kong to support the Tiananmen protest.</i>		
历史的伤口	zh.wikipedia.org/w & \$term	2 Talk:历史的伤口
<i>Wound of History, song title to support the Tiananmen protest, a collaboration by many Hong Kong and Taiwan musicians.</i>		
血染的風采	prefix.standard	1
<i>Blood-Stained Glory, a Chinese patriotic song, later used by Hong Kong people to commemorate the June 4th Incident.</i>		
人民不會忘記	zh.wikipedia.org/zh-hans/\$term	1 人民不會忘記
<i>The People Do Not Forget, a book by Hong Kong reporters who covered the Tiananmen protest in 1989.</i>		
自由花	prefix.standard	1
<i>Flower of Freedom, song title, inspired by China democracy movement. It became the theme song of Victoria Park June 4th Vigil.</i>		

Table 8.7: GFW Rulebook: Tiananmen

8.8 Miscellaneous

This section covers those terms not in any of the previous categories. Many terms here are surprising, amusing, puzzling and/or intriguing.

Term	GFW Rule	Pages Affected
Great_Wall_of_China	en.wikipedia.org/wiki/\$term	0-EN:8 Great_Wall_of_China(en) Great_Wall_of_China_hoax(en) Great_Wall_of_China_(album)(en) Great_Wall_of_China_Marathon(en) ...
<i>This one is very odd. The page is entirely non-sensitive. Why did GFW have this rule??</i>		
恶搞	prefix.standard zh.wikipedia.org/zh-cn/\$term	3 恶搞之家 恶搞文化
惡搞	prefix.standard zh.wikipedia.org/zh-cn/\$term	5 恶搞RFC 恶搞政黨 恶搞文化 恶搞校園
<i>The term literally means “evil fun”, similar to the notion “kuso” in Japanese. It usually involves parody.</i>		
kuso	prefix.standard	5 KUSO KUSO_KUSO 酷酷兽 KUSO_KUSO.酷酷獸 Kuso Kuso_game
<i>Wikipedia’s description is “(a) term used in East Asia for the internet culture that generally includes all types of camp and parody”.</i>		
火药	prefix.standard zh.wikipedia.org/zh-cn/\$term	6 火药塔 火药塔_(里加) 火药武器 火药池 火药阴谋
炸藥	prefix.standard	1
燃烧瓶	prefix.standard zh.wikipedia.org/zh-cn/\$term	1
燃燒彈	prefix.standard	1
燒夷彈	prefix.standard	1
<i>These five terms are gunpowder, explosives, petrol bomb/Molotov cocktail, incendiary bomb, napalm bomb, respectively.</i>		
俄羅斯	broad.standard	688 1917年俄羅斯革命 俄羅斯方塊 白俄羅斯 Talk:俄羅斯人 Template:俄羅斯歷史 Category:俄羅斯帝國 ...
<i>Russia, in traditional Chinese. This rule alone affects 688 Wikipedia pages.</i>		
观音法门	broad.standard	1
觀音法門	broad.standard	3 觀音法門_(新興宗教) Talk:觀音法門
<i>Quan Yin Method, a religion derived from buddhism.</i>		
Tokyo-Hot	zh.wikipedia.org/zh-cn/\$term	1 Tokyo-Hot
<i>a Japanese pornography brand.</i>		
大屠杀列表	prefix.standard	1
大屠殺列表	prefix.standard	1
<i>list of massacres.</i>		
獨立進行曲	prefix.standard	1
<i>Independence March, the page is about the Turkish national anthem.</i>		
2005年5月	prefix.standard	2 2005年5月逝世人物列表
<i>May 2005. It is a mystery why this term is on GFW rulebook. The author welcomes input regarding this term.</i>		
2012年人類滅亡説	prefix.standard	1
<i>2012 human extinction theory.</i>		
北京中民信房地产开发 有限公	zh.wikipedia.org/zh-hans/\$term	1 北京中民信房地产开发有限公司
<i>Beijing Central People’s Trust Real Estate Development Corporation Ltd, a company owned by Xi Jinping’s sister Qi Qiaoqiao and Qi’s husband Deng Jiagui, both names are on GFW rulebook.</i>		

Table 8.8: GFW Rulebook: Miscellaneous

8.9 Non-Articles

This section covers those rules specifically targeting non-article pages. Numbers in parentheses are counts of rules targeting that namespace.

- Namespace Project (7). These included debate page for page deletion, page protection, vandalism, and their archives. Apparently many subjects which caused these debates are “sensitive” ones. These debates can be fierce and shedding light not only on the subject itself, but also on who inclined to toe the Chinese Communist Party line among the participating editors.
- Namespace User (25). These 25 rules affect 10 users. For 7 of these, only a subpath of the User page is blocked; for the other 3, the entire User page is blocked.
- Namespace Category (2).
- Namespace File (2).
- Namespace Talk (23).

Most of these pages are obscure. Their inclusion in GFW seems to imply that China censorship officials and GFW operators have more-than-superficial knowledge of Wikipedia structure and content, which is quite contradictory to what we have seen so far. My speculation is that certain Chinese Wikipedia editors have interactions with Chinese censors and they provide these obscure pages for blocking. If it is the case I do not blame them, I think they acted out of good intention, hoping in return the China authority can lift the blanket ban and just block these specific pages.

Term	GFW Rule	Pages Affected
Namespace Project		
Wikipedia:頁面存廢討論	prefix.standard	2471 Wikipedia:頁面存廢討論/今天 Wikipedia:頁面存廢討論/修版本除 Wikipedia:頁面存廢討論/提報過程 ...
<i>This is the debate page for page deletion. This rule affects 2471 pages, most of which are archive pages.</i>		
Wikipedia:請求保護頁面/保護存檔	zh.wikipedia.org/zh-cn/\$term	37 Wikipedia:請求保護頁面/保護存檔 Wikipedia:請求保護頁面/保護存檔/2007年10月－12月 ...
<i>This is the archive for page protection requests.</i>		
Wikipedia:對話頁討論索引	prefix.standard	9 Wikipedia:對話頁討論索引/wikipedia Wikipedia:對話頁討論索引/主題 Wikipedia:對話頁討論索引/分類 ...
<i>This is an auto-generated page which lists “hot” Talk pages. Talk pages for sensitive topics are often hot due to edit wars.</i>		
Wikipedia:互助客棧/求助	prefix.standard	11 Wikipedia:互助客棧/求助/存檔/2007年1月 Wikipedia:互助客棧/求助/存檔/2006年12月 ...
<i>This is the page editors ask for help from each other.</i>		
Wikipedia:当前的破坏/存档/2010年7-	prefix.standard	1 Wikipedia:当前的破坏/存档/2010年7-12月
<i>This is vandalism archive for the period Jul to Dec 2010. Looking at the page content, quite several parts seem “sensitive”, but nothing stands out.</i>		
Wikipedia:刪除投票和請求/2007年6月	prefix.standard	30 Wikipedia:刪除投票和請求/2007年6月4日 Wikipedia:刪除投票和請求/2007年6月1日 ...
<i>Page deletion nomination and voting for Jun 2007. Like as discussed in Section 4.2, this rule is actually for “Wikipedia:刪除投票和請求/2007年6月4日” (for Jun 4, 2007). It gets truncated due to the 64-byte limit. That particular page contains a long debate on whether “六四18周年” (18th anniversary of June 4th Incident) should have its own article.</i>		
Wikipedia:聊天/2005年10月	prefix.standard	1
<i>This is the chat page for Oct 2005 and it is very worth reading. From the page content, we learned that Wikipedia was blocked in Oct 2005, and many Chinese Wikipedia editors were disappointed, some of whom blame that Chinese Wikipedia has “too much” political content. There are several long discussions on proposals to create a “clean” (self-censored) Chinese Wikipedia.</i>		
Namespace User		
User:Inspector/百度封禁	prefix.standard	1

continued on next page...

...continued from previous page		
Term	GFW Rule	Pages Affected
贴吧一览 <i>This page is about the list of topics forbidden by Baidu Tieba.</i>		
User:Lxr1234/535	prefix.standard zh.wikipedia.org/zh-hant/\$term	1
<i>This page contains the content of the June 4th Incident article.</i>		
User:Mongol/arch2	prefix.standard zh.wikipedia.org/zh-hk/\$term	1
<i>This is a personal archive page, I do not know why it is on GFW rulebook.</i>		
User:Mungs/香港人權狀況	prefix.standard zh.wikipedia.org/zh-hk/\$term zh.wikipedia.org/zh-sg/\$term zh.wikipedia.org/zh-hant/\$term	1
<i>This page is about Hong Kong human rights.</i>		
User:Qingmui	prefix.standard zh.wikipedia.org/zh-hk/\$term	1
<i>This user has the content of Charter 08 on his personal start page.</i>		
User:Tlrmq/六四紀念Temp	prefix.standard zh.wikipedia.org/zh/\$term zh.wikipedia.org/zh-cn/\$term zh.wikipedia.org/zh-hans/\$term	1
<i>This page has information about memorial events for the June 4th Incident.</i>		
User:Zhangjintao	prefix.standard zh.wikipedia.org/zh/\$term zh.wikipedia.org/zh-hans/\$term	54 User:Zhangjintao/MainPage User:Zhangjintao/Computers User:Zhangjintao/Favourites User:Zhangjintao/快速删除日誌 ...
<i>This user is a heavy Wikipedia editor who must have done something that offended certain people! His personal start page is currently empty.</i>		
User:感靈	prefix.standard	1
<i>This user's personal start page contains information about the June 4th Incident.</i>		
User:Philip/June.4	zh.wikipedia.org/zh-cn/\$term zh.wikipedia.org/zh-tw/\$term zh.wikipedia.org/zh-hk/\$term zh.wikipedia.org/zh-sg/\$term	1 User:PhiLiP/June.4
<i>This page redirects to the June 4th Incident article.</i>		
User:Liangent-bot/Base64URL/5YWt5Zub5LqL5Lu2	zh-cn/\$term zh.wikipedia.org/zh-hant/\$term zh.wikipedia.org/zh-hk & \$term	1 User:Liangent-bot/Base64URL/5YWt5Zub5LqL5Lu2 1 User:Liangent-bot/Base64URL/5YWt5Zub5LqL5Lu2
<i>This is a robot user; this page redirects to the June 4th Incident article. These three rules all look very odd, because the target string is long (more than b4 bytes), so GFW operators had to do some manual word here.</i>		
Namespace Category		
Category:政治書籍	prefix.standard Category:political books.	1
Category:中文書籍	zh.wikipedia.org/zh-tw/\$term Category:Chinese books.	1 Category:中文書籍
Namespace File		
File:The_Gate_of_Heavenly_Peace	prefix.standard This file is a picture of the Gate of Heavenly Peace, i.e. Tiananmen.	1 File:The_Gate_of_Heavenly_Peace.jpg
EpochTimes.svg	broad.standard This file is the logo of Epoch Times, a Falungong newspaper and website. This rule does not have the "File:" prefix, but it is apparent that the intention is to block this file.	1 File:EpochTimes.svg
Namespace Talk		
Talk:2013年《南方周末》	prefix.standard	2 Talk:2013年《南方周末》新年献词被删改事件 53 <i>continued on next page...</i>

...continued from previous page		
Term	GFW Rule	Pages Affected
新年献词		
<i>Talk page for one variation of the Southern Weekly Incident. Like as discussed in Section 4.2, this rule is actually for “Talk:2013年《南方周末》新年献词被删改事件”.</i>		
Talk:2013年《南方周末》新年献词	zh.wikipedia.org/zh-hant/\$term	2 Talk:2013年《南方周末》新年献词 Talk:2013年《南方周末》新年献词被删改事件
<i>Talk page for one variation of the Southern Weekly Incident. Like as discussed in Section 4.2, this rule is actually for “Talk:2013年《南方周末》新年献词”.</i>		
Talk:2013年《南方周末》新年特刊	prefix.standard	2 Talk:2013年《南方周末》新年特刊事件 Talk:2013年《南方周末》新年特刊被删改事件
<i>Talk page for one variation of the Southern Weekly Incident. Like as discussed in Section 4.2, this rule is actually for “Talk:2013年《南方周末》新年特刊被删改事件”.</i>		
Talk:陈光诚	prefix.standard	1
<i>Talk page for Chen Guangcheng, the legendary blind rights defending lawyer.</i>		
Talk:防火长城	prefix.standard	1
<i>Talk page for the Great Firewall. The main page title “防火长城” is not on GFW rulebook.</i>		
Talk:火药	prefix.standard	2 Talk:火药阴谋
<i>Talk page for gunpowder.</i>		
Talk:国际学校	prefix.standard zh.wikipedia.org/zh-tw/\$term zh.wikipedia.org/zh-hant/\$term	1 Talk:国际学校_(中国大陆)
<i>Talk page for international schools. The main article is not on GFW rulebook. It is mysterious that both the talk page and the main page (including the ones for international schools in mainland China) do not seem sensitive at all.</i>		
Talk:宋祖英	prefix.standard zh.wikipedia.org/zh/\$term zh.wikipedia.org/zh-cn/\$term zh.wikipedia.org/zh-tw/\$term zh.wikipedia.org/zh-hk/\$term zh.wikipedia.org/zh-sg/\$term zh.wikipedia.org/zh-hans/\$term	1
<i>Talk page for Song Zuying, singer, a household name in China. There is wide spread rumor about her relationship with Jiang Zemin. The main article is not blocked, even though it does cover that rumor. My hypothesis is that when the censors were studying this page, that paragraph did not exist on the main article. Apparently there has been edit wars over whether this rumor should be included. This also suggests that the Chinese censors have not revisited the main article for long.</i>		
Talk:太子党	prefix.standard	5 Talk:太子党/2005年存档 Talk:太子党/2006年存档 Talk:太子党/存档3 Talk:太子党/存档4
<i>Talk page for princelings.</i>		
Talk:瓮安骚乱	prefix.standard	1
<i>Talk page for the Weng'an Riot, a mass protest in Weng'an, Guizhou, on Jun 28, 2008. Surprisingly the main article is not on GFW rulebook.</i>		
Talk:西藏連環自焚事件	prefix.standard	1
<i>Talk page for the serial self-immolations in Tibet. The main article title is not on GFW rulebook, but a related term (藏區連環自焚事件) is.</i>		
Talk:一个中国	prefix.standard	1
<i>Talk page for “one China”, which is the official principle in China's dealing with Taiwan affairs. The main page is not on GFW rulebook, but the Talk page has a lot more negative content towards China authority.</i>		
Talk:中國茉莉花革命	prefix.standard	1
Talk:中國茉莉花活動	prefix.standard	0
<i>Talk pages for two terms about China's Jasmine Revolution. Both main articles are on GFW rulebook (by prefix rule).</i>		
Talk:左派和右派	prefix.standard	1
<i>Talk page for “leftist and rightist”. The main article is not on GFW rulebook.</i>		

Table 8.9: GFW Rulebook: Non-Articles

8.10 SELF

GFW rulebook contains some terms that are not targeting any specific websites, I call this type “SELF”. Excluding URL terms, this set is quite small. Out of this small set, 18 terms affect Wikipedia, which are listed in this section.

By intuition, people may think that these SELF terms must be super-sensitive, but in reality, some of them are not. I tend to believe that some of these terms are relics from old times, and some are due to lack of considerations.

Term	GFW Rule	Pages Affected
网络人权宣言	网络人权宣言	2 Talk:网络人权宣言
<i>Manifesto of Online Human Rights, a manifesto published by Chinese journalists, writers, scholars and lawyers on Oct 8, 2009.</i>		
自由亚洲电台	自由亚洲电台	2 Talk:自由亚洲电台
<i>Radio Free Asia.</i>		
Mao.The.Unknown.Story	Mao.The.Unknown.Story	2-EN:1 File:Mao.The.Unknown.Story.jpg File:Mao.The.Unknown.Story_(Paperback).jpg Mao.The.Unknown.Story(en)
<i>Book title. Its Chinese title in traditional Chinese 毛澤東：鮮為人知的故事 is blocked by the standard prefix rule.</i>		
天葬	天葬	3 天葬坟 Talk:天葬
<i>Sky burial, a traditional funeral ceremony in Tibet. It is also the title of a banned book on Tibetan affairs by Wang Lixiong.</i>		
我的奋斗	我的奋斗	3 Talk:我的奋斗 User:talk:我的奋斗1
<i>Mein Kampf, Adolf Hitler's autobiographical manifesto. I believe GFW operators used this term for testing in the very early stage of GFW.</i>		
延安日记	延安日记	2 Talk:延安日记
<i>Diary in Yan'an, a book by Soviet diplomat Peter Vladimirov, which covers the history of Chinese Communist Party from 1942 to 1945.</i>		
Ultrasurf	Ultrasurf	2-EN:4 File:UltraSurf.png Template:User_Ultrasurf Ultra-Surf(en) Ultrasurf(en) Talk:Ultrasurf(en) User:Ultrasurf2(en)
<i>A popular censorship-circumvention software. Its Chinese name is 无界, which is blocked by the standard prefix rule.</i>		
Ultrareach	Ultrareach	0-EN:1 Ultrareach(en)
<i>Ultrareach is the name of the company that develops Ultrasurf.</i>		
无界网络	无界网络	2 Talk:无界网络
<i>The full Chinese name of Ultrasurf. The short name 无界 is blocked by the standard prefix rule.</i>		
动态网	动态网	1
<i>Name of the company that develops the popular circumvention software Freegate (自由门). Its literal meaning is “dynamic web”.</i>		
盛雪	盛雪	6 盛雪_(臧錫紅) Talk:盛雪_(臧錫紅) User:Xiaogang_AU/盛雪 User:talk:盛雪 User:talk:青花盛雪
<i>Sheng Xue, dissident activist.</i>		
盘古乐队	盘古乐队	1
<i>Pangu band, an avant-garde punk band. The traditional version of a related term 盤古樂團 is also blocked, but not by itself.</i>		
王斌余	王斌余	1
<i>Wang Binyu, a migrant worker who committed homicide and was executed on Oct 19, 2005. Not many people know about or remember his case.</i>		
邓正来	邓正来	2 Talk:邓正来
<i>Deng Zhenglai, a scholar passed away on Jan 24, 2013. This rule has been around for long time though.</i>		
彭小枫	彭小枫	2 Talk:彭小枫
<i>Peng Xiaofeng, a high-ranking PLA general. The term's sensitivity is a mystery.</i>		
章沁生	章沁生	2 Talk:章沁生
<i>Zhang Qinsheng, a high-ranking PLA general. The term's sensitivity is a mystery.</i>		
6·4	6·4	0-EN:1 6·4(en)
<i>This is for the June 4th Incident. The character between 6 and 4 is a middle dot, a.k.a. interpunct or interpoint (unicode U+00B7).</i>		
平反六四	平反六四	5 平反六四議案 Talk:平反六四議案 Template:User_平反六四 Category:要求平反六四的维基人
<i>redress June 4th.</i>		

Table 8.10: GFW Rulebook: SELF

8.11 URLs

GFW rulebook contains a large number of URLs, in the order of 10^3 to 10^4 , and many of them affect Wikipedia, especially the English Wikipedia's non-article namespaces like Users. There are too many to include in this report, so I only report two types: 1. those affect pages in the Chinese Wikipedia, or 2. those affect articles (i.e. Namespace = 0) in the English Wikipedia. For the latter type, the third column of the table "Pages Affected" only considers articles, unlike the previous sections.

Note that in GFW rulebook, many URL terms require the dot before the website name, e.g. the leading dot in ".businessweek.com". Many Wikipedia articles do not have that dot in its page title, thus avoid offending GFW, (hence not included in the table below).

The table include several IP addresses, and some domains which are abandoned.

Term	GFW Rule	Pages Affected
URLs Affecting Chinese Wikipedia		
67.15.34.251	67.15.34.251	1 User:67.15.34.251
123.204.163.6	123.204.163.6	1 User.talk:123.204.163.60/存档
204.12.226.163	204.12.226.163	1 User.talk:204.12.226.163/存档
204.74.211.115	204.74.211.115	1 User.talk:204.74.211.115/存档
.clrc.cc	.clrc.cc	1 User.talk:Www.clrc.cc
.isiwa.cn	.isiwa.cn	2 User:Www.isiwa.cn User.talk:Www.isiwa.cn
qian.li	qian.li	1 User.talk:Qian.lian
URLs Affecting English Wikipedia Articles		
64tianwang.com	64tianwang.com	0-EN:1 64tianwang.com(en)
adultfriendfinder.com	adultfriendfinder.com	0-EN:1 Adultfriendfinder.com(en)
asianews.it	asianews.it	0-EN:1 Asianews.it(en)
blogspot.com	blogspot.com	0-EN:3 .blogspot.com(en) Blogspot.com(en) Cakewrecks.blogspot.com(en)
.businessweek.com	.businessweek.com	0-EN:1 Investing.businessweek.com(en)
crackle.com	crackle.com	0-EN:1 Crackle.com(en)
cultdeadcow.com	cultdeadcow.com	0-EN:1 Cultdeadcow.com(en)
epochtimes.com	epochtimes.com	0-EN:1 Epochtimes.com(en)
.facebook.com	.facebook.com	0-EN:1 Www.facebook.com(en)
favstar.fm	favstar.fm	0-EN:1 Favstar.fm(en)
fleshbot.com	fleshbot.com	0-EN:1 Fleshbot.com(en)
.fulltiltpoker.com	.fulltiltpoker.com	0-EN:1 Www.fulltiltpoker.com(en)
ishr.org	ishr.org	0-EN:1 Ishr.org(en)
ladbrokes.com	ladbrokes.com	0-EN:1 Ladbrokes.com_Championship(en)
megavideo.com	megavideo.com	0-EN:1 Megavideo.com(en)
njuice.com	njuice.com	0-EN:1 ToonJuice.com(en)
.nrk.no	.nrk.no	0-EN:1 Www.nrk.no(en)
.nytimes.com	.nytimes.com	0-EN:1 Www.nytimes.com(en)
pastebin.com	pastebin.com	0-EN:1 Pastebin.com(en)
phayul.com	phayul.com	0-EN:1 Phayul.com(en)
radiobeta.com	radiobeta.com	0-EN:1 Radiobeta.com(en)
rebuildhk.com	rebuildhk.com	0-EN:1 Rebuildhk.com(en)
tinychat.com	tinychat.com	0-EN:1 Tinychat.com(en)
twitoaster.com	twitoaster.com	0-EN:1 Twitoaster.com(en)
twitpic.com	twitpic.com	0-EN:1 Twitpic.com(en)
twitterfall.com	twitterfall.com	0-EN:1 Twitterfall.com(en)
veoh.com	veoh.com	0-EN:1 Veoh.com(en)
.youporn.com	.youporn.com	0-EN:1 Www.youporn.com(en)
.youtube.com	.youtube.com	0-EN:1 Www.youtube.com(en)
youtu.be	youtu.be	0-EN:1 Youtu.be(en)

Table 8.11: GFW Rulebook: URLs

9 Conclusion Remarks

In this study, we examined the entire Wikipedia corpus (Chinese version and English version) and revealed the complete and exact GFW rulebook for Wikipedia. In addition, we examined GFW's HTTP response filtering scheme thoroughly and surfaced a small but comprehensive list. A sample of notable findings are:

- There are 78 terms for which GFW blocks a non-standard variant but not the canonical path. These are cases the censors intend to block but the block does not really happen, suggesting the censors have poor understanding of Wikipedia's content and serving system.
- Many obscure non-article pages are blocked, which raises suspicion that these pages were provided to the censorship bureaucrats by Wikipedia editors who are very familiar with the content (e.g. those who participated in the edit wars and/or discussions regarding self-censorship proposals).
- GFW string matching rules have a 64-byte hard limit of size.
- GFW's HTTP request filtering and HTTP response filtering are two separate systems. The latter has a lot more heterogeneity.

The biggest learning out of this study, in my opinion, is that GFW operation is haphazard and ill-maintained. Also, there are many indications that the GFW operators are somewhat disconnected from the censorship bureaucrats.

We hope the revealing can be of interest to internet censorship watchers, Wikipedia researchers, China observers, and ordinary Chinese citizens.

9.1 Future Work

- The methodology in this study can be applied to more than just Wikipedia.
- Naturally, it is nice to have a monitoring system to track changes of GFW's filtering rules. Using the methodology in this report, the monitoring can be much more efficient than checking individual Wikipedia URLs like Greatfire.org's current approach.
- Given our research result on GFW's HTTP response filtering, we learn that the response filtering happens in a more distributed fashion and different ISPs may have deployed different filtering rules. By examining which rules are effective in which regions, we will be able to learn more on GFW topology.
- The other type of work is from the social and political perspective. Because the list reported in this document is *complete*, we know not only about what topics are censored, but also what topics are *not* censored. This powerful comparison will give us deep insights about the China censors, e.g., what is their priority.

9.2 Acknowledgements

This research is a solo project over a short period of time, so this is a short list. The author thanks Jedidiah Crandall et al. for their excellent ConceptDoppler paper [5], and the operators of the Greatfire.org site for diligently tracking GFW and China censorship. Furthermore, I do like to mention Dr Xu Zhiyong ([English Wikipedia](#), [Chinese Wikipedia](#)), the iconic figure of [China's New Citizen Movement](#). His passion, courage and coolness are inspiration to the author, and in particular, the fact that his Wikipedia entry is currently *not* blocked motivated me to conduct a thorough study of GFW's blacklist for Wikipedia. Indeed, China censors do intend to block the Xu Zhiyong article, but it only blocks a non-standard variant.

References

- [1] Greatfire.org. <https://www.greatfire.org> (since 2011).
- [2] Citizen Lab. <https://china-chats.net> (since Jul 2013).
- [3] Crandall, J. et al. (2013). Chat program censorship and surveillance in China: Tracking TOM-Skype and Sina UC, <http://firstmonday.org/ojs/index.php/fm/article/view/4628/3727>.
- [4] China Digital Times. <http://chinadigitaltimes.net/chinese/category/网情透视/敏感词库/> (since Apr 2011).
- [5] Crandall, J. et al. (2007). ConceptDoppler: A Weather Tracker for Internet Censorship http://www.cs.unm.edu/crandall/concept_doppler_ccs07.pdf
- [6] Zittrain, J. and Edelman, B. (2002). Empirical Analysis of Internet Filtering in China.
- [7] Global Internet Freedom Consortium (2002). The Great Firewall Revealed, <http://www.internetfreedom.org/files/WhitePaper/ChinaGreatFirewallRevealed.pdf>.

Appendices

A Diagnosis of Greatfire.org's Wikipedia List

Greatfire.org keeps track of status of ~ 700 Wikipedia pages, out of which 393 pages are claimed to be 100% blocked or partially blocked. I examined these pages thoroughly. The following is the result:

- 240 pages offend GFW's rulebook of HTTP request scan, so they are not accessible from China. A user trying to load these pages will immediately get an error page in the browser. In a probing session, we will get an instant connection failure.
- 14 articles are accessible from China. They were tested from more than 20 different proxy IPs on 2013/10/13 and 2013/10/17. For these pages, Greatfire's "blockage percentage" ranges from 10% to 56%. This percentage is averaged over multiple tests. There can be a few explanations, for example, it could be that there were offensive terms in the page content in the past, but is no longer there now, or there is a certain level of noise in Greatfire.org's testing.
- 139 articles are interrupted by GFW. A user can generally load part of the page, but it would hang and/or get reset after a while. We examined thoroughly these pages by probing, and they provide a very valuable source for our study of GFW's HTTP response filtering (Section 5).

I put these results on a Google spreadsheet at <http://goo.gl/jUJpHb>. This spreadsheet contains all entries that are reported as blocked or partially blocked by Greatfire.org. They are grouped into two sheets. The first contains those do not offend GFW's HTTP request scan and the second contains those do. The order on each sheet is Greatfire.org's default order, except for the 14 accessible articles, which are put at the bottom of the first sheet. For those inaccessible pages due to GFW's HTTP response filtering, we list the offending string as well. In cases where the page contains multiple offending strings, we list the first occurrence.

B Self Censorship Efforts on Chinese Wikipedia

During this study, we encountered many records which indicate self censorship efforts on Chinese Wikipedia. An excellent feature of Wikipedia is that all edits are recorded and most can be retrieved, so Wikipedia become an extremely valuable resource to study the cultural phenomenon of self censorship in China.

According to the Washington Post news article "[Reference Tool On Web Finds Fans, Censors](#)" (Feb 20, 2006), "*Wikipedia received positive coverage in China's state press in early 2004, but it was blocked on 3 June 2004... Proposals to practice self-censorship in a bid to restore the site were rejected by the Chinese Wikipedia community.*" In our study, we found records of such proposals, e.g. [Wikipedia:聊天/2005年](#) (chat archive of 2005), also another relevant page: [Wikipedia:删除投票和请求/2007年6月4日](#) (page delete proposal archive on June 4th, 2007). The debates were lengthy and heated. Both pages are intentionally blocked by GFW.

We see proposals like:

- *Maybe we can make Chinese Wikipedia an "apolitical" encyclopedia, e.g., empty all sensitive articles and lock them up.*
- *Wikipedia should obey local laws. We should monitor and modify content which break laws, thus avoid being blocked.*
- *Stick to Wikipedia spirit, cooperate with the authority.* (sounds a bit funny, no?)

From the discussion records, we can see that this is not a minority sentiment. There can be three types of users who demand self censorship:

- A: Those who believe in Wikipedia spirit, but feel a compromise may work better for China in the long run.
- B: Those who believe in the Party and voluntarily toe the Party line, due to their education/China's propaganda.
- C: Those who are affiliated or sponsored by the China authority. I can not find evidence for this type from cursory browsing though.

After the lengthy blockage and the emergence of Baidu Baike (Baidu's copycat of Wikipedia), many Type B editors migrate to Baidu Baike. But after the lift of the ban, some Type B editors returned. Anecdotes show that there are more heated edit wars for sensitive articles not blocked by GFW, than those blocked ones. As a result, these unblocked articles become more favorable to the China authority than otherwise.

The final note is that my gut feeling is certain Chinese Wikipedia editors have helped China censors by providing list of sensitive pages for blocking. The basis of this speculation is the existence of those obscure Wikipedia non-article pages on GFW rulebook (Section 8.9). I am an adamant opponent to GFW and internet censorship, but I feel a Wikipedia with most of its content accessible to Chinese users and only a small number of pages blocked, could still be net positive for China.